



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

BEST AVAILABLE COPY

REC'D 03 SEP 2004

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03013957.0

PRIORITY  
DOCUMENT

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

R C van Dijk



Anmeldung Nr:  
Application no.: 03013957.0  
Demande no:

Anmeldetag:  
Date of filing: 20.06.03  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER  
WISSENSCHAFTEN E.V.

14195 Berlin  
ALLEMAGNE  
MAX-DELBRÜCK-CENTRUM FÜR MOLEKULARE MEDIZIN  
Robert-Rössle-Strasse 10  
13125 Berlin  
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se referer à la description.)

Disease related protein network

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)  
revendiquée(s)  
Staat/Tag/Aktenzeichen/State>Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

G01N33/48

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL  
PT RO SE SI SK TR LI

## DISEASE RELATED PROTEIN NETWORK

The present invention relates to a method for generating a network of direct and indirect interaction partners of a disease-related (poly)peptide comprising the steps of (a) contacting a selection of (poly)peptides suspected to contain one or several of said direct or indirect interaction partners with said disease-related (poly)peptides and optionally with known direct or indirect interaction partners of said disease-related (poly)peptide under conditions that allow the interaction between interaction partners to occur; (b) detecting (poly)peptides that interact with said disease-related (poly)peptide or with said known direct or indirect interaction partners of said disease-related (poly)peptide; (c) contacting (poly)peptides detected in step (b) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (b) under conditions that allow the interaction between interaction partners to occur; (d) detecting proteins that interact with said (poly)peptides detected in step (b); (e) contacting said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide, said (poly)peptides detected in steps (b) and (d) and a selection of proteins suspected to contain one or several (poly)peptides interacting with any of the afore mentioned (poly)peptides under conditions that allow the interaction between interaction partners to occur; (f) detecting (poly)peptides that interact with said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide or with said (poly)peptides identified in step (b) or (d); and (g) generating a (poly)peptide - (poly)peptide interaction network of said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide and said (poly)peptides identified in steps (b), (d) and (f). Moreover, the present invention relates to a protein complex comprising at least two proteins and to methods for identifying compounds interfering with an interaction of said proteins. Finally, the present invention relates to a pharmaceutical composition and to the use of compounds identified by the present invention for the preparation of a pharmaceutical composition for the treatment of Huntington's disease.

Several documents are cited throughout the text of this specification. The disclosure content of the documents cited herein (including any manufacturer's specifications, instructions, etc.) is herewith incorporated by reference.

With the identification of >35,000 genes in the human genome the challenge arises to assign biological function to all proteins and to link these proteins to physiological pathways and disease processes. Since protein-protein interactions play a role in most events in a cell, clues to the function of an unknown protein can be obtained by investigating its interaction with other proteins whose function are already known. Thus, if the function of one protein is known, the function of the binding partners can be inferred (deduced). This allows the researcher to assign a biological function to uncharacterized proteins by identifying protein-protein interactions. For example, several so far uncharacterized proteins in *Caenorhabditis elegans* were identified in a yeast two-hybrid screen for eukaryotic 26S proteasome interacting proteins and thereby could be linked to the ubiquitin-proteasome proteolytic pathway (Vidal et al., 2001). Elucidation of protein-protein interactions is particularly desired when it comes to the generation of new drugs. For many diseases, the available drug portfolio is insufficient or inappropriate to provide a cure or to prevent onset of the disease. One such disease is Huntington's disease.

Huntington's disease (HD) is a neurodegenerative disorder caused by an expanded polyglutamine (polyQ) tract in the multidomain protein huntingtin (htt). The elongated polyQ sequence is believed to confer a toxic gain of function to htt. It leads to htt aggregation primarily in neurons of the striatum and cortex and subsequently to the appearance of the disease phenotype. However, there is experimental evidence that loss of htt function may also contribute to HD pathogenesis. Since huntingtin aggregation correlates with disease progression, it is crucial to develop methods for identifying factors that promote or inhibit aggregation of huntingtin.

Previously, a number of single interaction partners of huntingtin had been reported. In light of these reports, it is tempting to speculate that huntingtin is bound into a larger network of interacting partners, many of which might be capable of modulating huntingtin's activity and function by direct or indirect interaction. It is likely that an

aberrant interaction of huntingtin with some of the members of said network will impair huntingtin's normal function. Moreover, this interaction might also be relevant for the conformation of huntingtin or for its solubility or state of aggregation. Interfering with the direct or indirect interactions of the protein-protein interaction network will provide an excellent basis for therapeutic intervention as it will allow to modulate huntingtin's activity or state of aggregation or both. The state of the art so far did not provide compounds capable of reducing or suppressing huntingtin aggregation since the factors promoting or suppressing huntingtin aggregation were not known.

Thus, the technical problem underlying the present invention was to provide novel approaches for identifying direct or indirect interaction partners of disease-related proteins, which must be seen as new targets for drug development. The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a method for generating a network of direct and indirect interaction partners of a disease-related (poly)peptide comprising the steps of (a) contacting a selection of (poly)peptides suspected to contain one or several of said direct or indirect interaction partners with said disease-related (poly)peptides and optionally with known direct or indirect interaction partners of said disease-related (poly)peptide under conditions that allow the interaction between interaction partners to occur; (b) detecting (poly)peptides that interact with said disease-related (poly)peptide or with said known direct or indirect interaction partners of said disease-related (poly)peptide;(c) contacting (poly)peptides detected in step (b) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (b) under conditions that allow the interaction between interaction partners to occur; (d) detecting proteins that interact with said (poly)peptides detected in step (b); (e) contacting said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide, said (poly)peptides detected in steps (b) and (d) and a selection of proteins suspected to contain one or several (poly)peptides interacting with any of the afore mentioned (poly)peptides under conditions that allow the interaction between interaction partners to occur; (f) detecting (poly)peptides that interact with said disease-related (poly)peptide and

optionally said known direct or indirect interaction partners of said disease-related (poly)peptide or with said (poly)peptides identified in step (b) or (d); and (g) generating a (poly)peptide-(poly)peptide interaction network of said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide and said (poly)peptides identified in steps (b), (d) and (f).

In accordance with the present invention, the term "direct and indirect interaction partners" relates to (poly)peptides that either directly interact with the disease-related (poly)peptide (direct interaction) or that interact via a protein binding to/interacting with said disease-related (poly)peptide. In the latter case, there is no direct contact between the direct interaction partner and the disease-related protein. Rather, a further protein forms a "bridge" between these two proteins.

The term "known direct or indirect interaction partners" refers to the fact that for certain disease-related (poly)peptides, such interaction partners are known in the art. If such interaction partners are known in the art, it is advantageous to include them into the method of the invention. If no such interaction partners are known in the art, then the network may be generated starting solely from the known disease-related (poly)peptide.

The term "conditions that allow the interaction between interaction partners to occur" relates to conditions that would, as a rule, resemble physiological conditions. Conditions that allow protein actions are well known in the art and, can be taken, for example from Golemis, E.A. Ed., Protein-Protein Interactions, Cold Spring Harbor Laboratory Press, 2002.

The term "suspected to contain one or more of said direct or indirect interaction partners" relates to the fact that normally, a selection of (poly)peptides would be employed where the person skilled in the art would expect that interaction partners are present. Examples of such selections of (poly)peptides are libraries of human origin such as cDNA libraries or genomic libraries.

The term "detecting proteins" refers to the fact that the (poly)peptides interacting with the "bait" (poly)peptides are identified within the selection of (poly)peptides. A further characterization or isolation of the "prey" (poly)peptides at this stage may be

advantageous but is not necessary. The term "detecting (poly)peptides" preferably also comprises characterizing said (poly)peptides or the nucleic acid molecules encoding said (poly)peptides. The skilled person knows that this can be done by a number of techniques, some of which are described for example in Sambrook et al., "Molecular Cloning, A Laboratory Manual"; CSH Press, Cold Spring Harbor, 1989 or Higgins and Hames (eds.). For example, the nucleotide sequence may be determined by DNA Sequencing, including PCR-Sequencing (see for example Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H., Cold Spring Harb Symp Quant Biol. 1986;51 Pt 1:263-73). Alternatively, the amino acid sequence of said (poly)peptide may be determined. The skilled artesian knows various methods for sequencing proteins which include the method of Edman degradation, which is a preferred method of the present invention of determining the amino acid sequence of a protein. However, the amino acid sequence of a protein or (poly)peptide can also be reliably determined by methods such as for example Maldi-Tof, optionally in combination with the method of Edman degradation. The interaction partner may be identified either as fusion with a DNA binding domain or as fusion with an activation domain. Preferably, if an interaction partner has been identified as a fusion molecule comprising a DNA binding domain, the interaction partner is cloned into a vector allowing the expression of the interaction partner as a fusion with an activation domain. Consequently, protein interaction can be tested in the context the DNA activation or the DNA binding domain.

In accordance with the present invention, the first round of detecting (poly)peptides that interact with the "bait" (poly)peptides recited in step (a) wherein the detected (poly)peptides be considered as "prey" (poly)peptides is followed by the second round of detecting further interacting (poly)peptides wherein the former "prey" (poly)peptides are now used as "bait" (poly)peptides. In certain preferred embodiments of the present invention such as in a two-hybrid detection system, a re-cloning of the former "prey" (poly)peptides into vectors that are suitable for expressing "bait" (poly)peptides may be desired.

Accordingly, the invention describes a novel strategy to identify protein-protein interaction networks for human disease proteins. This strategy was applied to detect pair-wise protein-protein interactions for Huntington's disease and is useful for other

hereditary diseases as well. Several human hereditary diseases are summarized in table 5.

A crucial step of the method of the invention is step (e). Here, the disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide are contacted under appropriate conditions, preferably at the same time, with both the (poly)peptides identified in steps (b) and (d) and further with a selection of (poly)peptides suspected to contain further interaction partners. Alternatively, the various baits, preys and further selection partners are added one after another, so that the final pool contains all baits and preys so far identified, in addition to the further selection partners. In other terms, in this step of the method of the invention, all "baits" and all "preys" are pooled and, additionally, further potential interaction partners are added. In this way, surprisingly the number of directed or indirect interactions partners of the previously identified "baits" and "preys" could significantly be enhanced. It is to be understood that various preys identified in one detection step may interact with each other and not only with the baits that were employed for the identification. For example, if a collection of baits detects preys "a" and "b", the invention does not exclude that "a" also interacts with "b". The same holds true mutatis mutandis for the baits used in accordance with the present invention.

It is further preferred in accordance with the present invention that the interaction of proteins is a specific interaction, such as a specific binding. This means that the (poly)peptide being an interaction partner with a further (poly)peptide only or essentially only interacts with the interaction site(s) involved with this interaction partner. This does not exclude, of course, that further interaction sites of said (poly)peptide interact with further interaction partners, wherein in the corresponding interaction is preferably also specific. The concept also embraces that, if a (poly)peptide has several identical interaction sites, which in nature bind to different interaction partners, these different interaction partners are also bound by the (poly)peptide in the method of the present invention.

In other terms, at least in the case of huntingtin, the number of interaction partners found in step (e) was enhanced in an exponential rather than in a linear fashion.

The term "(poly)peptide" refers alternatively to peptide or to (poly)peptides. Peptides conventionally are covalently linked amino acids of up to 30 residues, whereas polypeptides (also referred to as "proteins") comprise 31 and more amino acid residues.

The term "huntingtin" refers to a protein with the data bank accession number P42858 which is referenced for the purpose of the present invention as "wild-type huntingtin protein". However, the term "huntingtin" also comprises proteins encoded by the nucleic acid sequence deposited under accession number L12392 or to proteins encoded by nucleic acid molecules which hybridize to the nucleic acid molecule of L12392 under stringent conditions of hybridization. The present invention relates to all variants of the huntingtin protein. In particular, relevant for the present invention are those variants of huntingtin which comprise a polyglutamine tract (polyQ tract) or an elongated polyQ tract. A polyQ tract consists of two or more glutamines within the huntingtin protein. The insertion of additional glutamine codons will result in huntingtin proteins with, for example 2, 51, 75 or 100 added glutamines in comparison to the sequence deposited under accession number P42858. In fact, the person skilled in the art knows that the huntingtin protein may have a glutamine tract with any random number of glutamines in the range of 1 to 200 added glutamines. All these proteins are comprised by the present invention.

The term "hybridizes under stringent conditions", as used in the description of the present invention, is well known to the skilled artisan and corresponds to conditions of high stringency. Appropriate stringent hybridization conditions for each sequence may be established by a person skilled in the art on well-known parameters such as temperature, composition of the nucleic acid molecules, salt conditions etc.; see, for example, Sambrook et al., "Molecular Cloning, A Laboratory Manual"; CSH Press, Cold Spring Harbor, 1989 or Higgins and Hames (eds.), "Nucleic acid hybridization, a practical approach", IRL Press, Oxford 1985, see in particular the chapter "Hybridization Strategy" by Britten & Davidson, 3 to 15. Stringent hybridization conditions are, for example, conditions comprising overnight incubation at 42° C in a solution comprising: 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by

washing the filters in 0.1x SSC at about 65°. Other stringent hybridization conditions are for example 0.2 x SSC (0.03 M NaCl, 0.003M Natriumcitrat, pH 7) bei 65°C. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC). Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

The skilled person knows that the presence of additional codons in the nucleic acid sequence of huntingtin might significantly reduce the capability of this nucleic acid molecule to hybridize to the nucleic acid molecule deposited under L12392 and referenced as wild-type huntingtin protein. Nevertheless, such proteins shall still be comprised by the present invention. In fact, computer programs such as the computer program Bestfit (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) or blast, capable of calculating homologies between two nucleic acid sequences, efficiently recognize nucleotide insertions and allow for an adjustment of gaps created by these insertions. The term "huntingtin" as used in the present invention, also includes those molecules of huntingtin, which have a homology of more than 95% to wild-type huntingtin when analyzed with a program like bestfit under conditions not weighing gaps created by polyQ tracts (gap penalty=0).

The term "contacting" means bringing into contact so that two or more proteins or (poly)peptides can interact with each other, preferably under physiological conditions. The terms "interacting" or "binding" refer to a transient or permanent contact between two proteins or (poly)peptides. Preferably, the (poly)peptide or protein is provided by expression from a nucleic acid molecule, more preferably from a cDNA molecule within a cDNA library. Alternatively, said nucleic acid molecule is a genomic nucleic acid molecule of a genomic DNA library, or a nucleic acid molecule from a synthetic

DNA or RNA library. Preferably, the nucleic acid molecule encoding the disease-related protein or its interaction partner is obtainable from nerve cells, brain tissue human adrenal gland, human bladder, human bone, human brain, human colon, human dorsal root ganglion, human heart, human HeLa cells, human kidney, human liver, human lung, human mammary gland, human ovary, human pancreas, human placenta, human prostate, human retina, human salivary gland, human skeletal muscle, human small intestine, human smooth muscle, human spinal cord, human spleen, human stomach, human testis, human thymus, human thyroid, human tonsil, human trachea, human uterus, human cell line HEP G2, human cell line MDA 435, human fetal brain, human fetal heart, human fetal kidney, human fetal liver, human fetal spleen, human fetal thymus, human breast tumor, human cervix tumor, human colon tumor, human kidney tumor, human lung tumor, human ovary tumor, human stomach tumor, human brain tumor and/or human uterus tumor.

The term "disease-related protein" refers to a protein known to be the causative agent of a disease or known to be involved in onset or progression of a disease. Preferably, said disease is CHOREA HUNTINGTON or the disease-related protein is huntingtin. More preferably, the disease-related protein is selected from table 1 and/or 2. The term "conditions that allow the interaction between interaction partners" means conditions that are similar to physiological conditions. Preferably, said conditions are physiological conditions.

The term "selection of (poly)peptides" refers to a library of (poly)peptides which comprises the above-mentioned libraries, but also includes libraries such as phage display libraries. Preferably, the (poly)peptide is provided by expression from a nucleic acid molecule. Preferably, the protein or (poly)peptide expressed by said nucleic acid molecule is a (poly)peptide comprising a DNA binding domain (DBD) (in this case the fusion protein is termed "bait") or (b) a (poly)peptide comprising an activation domain capable of interacting with a transcription factor or an RNA polymerase and capable of activating transcription of a reporter or indicator gene (in this case the fusion protein is called "prey"). As used here, the terms "reporter gene" and "indicator gene" are to be understood as synonyms. It is important to note that one of the interaction partners will always comprise the amino acid sequence of a protein or (poly)peptide translated from said nucleic acid molecule while the other

interaction partner will comprise the amino acid sequence of a protein or protein fragment. Preferably, a bait used for a method of the present invention is selected from the proteins listed in table 1 and/or 2. If, for example, the proteins encoded by the nucleic acid molecules contain a DNA binding domain fused in frame, the fusion protein can bind to the DNA recognition sequence of the DNA binding domain. Interaction of said fusion protein with a second fusion protein containing an activation domain can induce transcription of a nearby indicator gene. The indicator gene may encode a selection marker such as a protein that confers resistance to an antibiotic including ampicillin, kanamycin, chloramphenicol, tetracyclin, hygromycin, neomycin or methotrexate. Further examples of antibiotics are Penicillins: Ampicillin HCl, Ampicillin Na, Amoxycillin Na, Carbenicillin disodium, Penicillin G, Cephalosporins, Cefotaxim Na, Cefalexin HCl, Vancomycin, Cycloserine. Other examples include Bacteriostatic Inhibitors such as: Chloramphenicol, Erythromycin, Lincomycin, Tetracyclin, Spectinomycin sulfate, Clindamycin HCl, Chlortetracycline HCl. Additional examples are proteins that allow selection with Bacteriosidal inhibitors such as those affecting protein synthesis irreversibly causing cell death. Aminoglycosides can be inactivated by enzymes such as NPT II which phosphorylates 3'-OH present on kanamycin, thus inactivating this antibiotic. Some aminoglycoside modifying enzymes acetylate the compound and block their entry in to the cell. Gentamycin, Hygromycin B, Kanamycin, Neomycin, Streptomycin, G418, Tobramycin Nucleic Acid Metabolism Inhibitors, Rifampicin, Mitomycin C, Nalidixic acid, Doxorubicin HCl, 5-Flurouracil, 6-Mercaptopurine, Antimetabolites, Miconazole, Trimethoprim, Methotrexate, Metronidazole, Sulfametoxazole. Alternatively, said indicator gene may encode a protein such as lacZ, GFP or luciferase, the expression of which can be monitored by detection of a specific color. Other proteins commonly used as indicator proteins are beta-galactosidase, beta-glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). In general, however, the selection in the yeast two hybrid-system is based on a deficiency of the yeast strain to produce specific amino acids. The skilled person knows that any amino acid deficiency can be used for this selection strategy.

Preferably said preys and baits are expressed from two separate expression vectors contained in one host cell. The nucleic acid molecule encoding the preys and baits can be introduced into the host cell, for example, by transformation, transfection, transduction or microinjection which are common techniques known to the person skilled in the art and which require no additional explanation. In addition, the nucleic acid molecule contains a chromosomal or episomal nucleic acid sequence encoding the above-mentioned indicator protein. The expression of said indicator protein is under control of a recognition sequence which serves as a binding site for the bait protein. The nucleic acid molecule may be fused either to a DNA binding domain or to an activation domain. Co-expression of only those bait- and prey fusion proteins which are capable of interacting will induce the expression of one of the above-identified indicator proteins and thus allow the identification a nucleic acid molecule encoding a protein capable of interacting with huntingtin or an interaction or binding partner of huntingtin. The skilled person knows this system as the yeast two hybrid system. The yeast two hybrid system, which uses a bait protein-prey protein combination to induce transcription of the reporter gene, is a preferred method to identify proteins capable of interacting with huntingtin or with a direct or indirect interaction or binding partner of huntingtin. See for example Fields and Song, Nature 340:245 (1989) or Uetz et al., 2000 Nature 403(6770): 623-7. This is a useful way of determining protein-protein interactions. Another preferred method uses the yeast three hybrid system, as described in U.S. Pat. No. 5,928,868. Preferably, steps (a) to (d) of the method for generating a network of direct and indirect interaction partners comprise the yeast two hybrid system. Preferably, steps (e) and (f) of the method for generating a network of direct and indirect interaction partners comprise yeast interaction mating. Preferably, said "interaction mating" comprises the interaction of all interaction partners identified in steps (a) to (d). Also preferred is that the interaction partners identified in steps (a) to (d) interact as prey and bait proteins, so that all prey proteins are contacted with all bait proteins. Using the array mating system, each bait is tested individually for interaction with every prey in the array. Alternatively, steps (e) and (f) of the method for generating a network of direct and indirect interaction partners comprise testing all interaction partners identified in steps (a) to (d) in interaction assays such as biacore or coimmunoprecipitation. When performing such an assay, it is preferred that the interaction partners are tested as prey and/or bait fusion proteins or contain no fused (poly)peptides. Preferably, all

interaction partners are contacted in the biacore or coimmunoprecipitation assay by themselves and by all other remaining interaction partners identified in steps (a) to (d).

The method for generating a network of direct and indirect interaction partners of a disease related protein or (poly)peptide has proven to be an effective tool for unveiling the protein-protein interactions (PPI) of preferably monogenic diseases. This is exemplified by the analysis of the disease related protein of Chorea Huntington, the analysis of which has demonstrated that the method of the present invention will be useful in an approach to identify potential drugs in the treatment of CHOREA HUNTINGTON. Moreover, this method will also be effective in unveiling the protein-protein interactions of other disease related proteins and in identifying novel targets for treatment of these diseases. Using a preferred combination of library and matrix yeast two-hybrid screens, based on the methods of the present invention, a highly connected network was generated among 70 proteins involved in 117 protein-protein interactions, 99 of which had not been described previously. As progression of Huntington's disease (HD) appears to be linked to huntingtin aggregation, a set of network proteins was tested for their potential to modulate this process. By using the methods of the present invention, it was discovered that the GTPase activating protein GIT1 strongly promotes huntingtin aggregation *in vivo*. GIT1 also localises to huntingtin aggregates in brains of transgenic mice and HD patients. Therefore, a combination of the methods of the present invention has proven to provide effective means for the identification of potential targets for therapeutic intervention. GIT1 is a selected example of a modulator interaction partner of huntingtin. The other proteins in the network of interaction partners disclosed by the present invention are further modulator interaction partners of huntingtin.

Preferably, the interaction mating comprises using an array mating system. In general, for this screen, MAT $\alpha$  yeast cultures are transformed with plasmids encoding prey proteins and arrayed on a microtiter plate for interaction mating with individual MAT $\alpha$  strains expressing bait proteins. Using this test system, each bait can be tested individually for interaction with every prey in the array. Diploid yeast clones, formed by mating on YPD plates and expressing both, bait and prey

proteins, are selected on agar SDII plates, and further transferred for example by a spotting robot on SDIV plates to select for protein-protein interactions. In a more preferred embodiment of the method, plasmids encoding bait and prey proteins are transformed into strains L40ccua and L40cca, respectively. L40cca clones are arrayed on microtitre plates and mixed with a single L40ccua clone for interaction mating. These cells are transferred, preferably by a robot onto YPD medium plates and, after incubation for 20h to 28h at approximately 30°C, for selection of the cells, were transferred onto SDII medium plates, where mating takes place, for additional 60h to 80h at approximately 30°C. For two-hybrid selection diploid cells are transferred onto SDIV medium plates with and without nylon or nitrocellulose membranes and incubated for approximately 5 days at about 30°C. The nylon or nitrocellulose membranes are subjected to the  $\beta$ -GAL assay. Positive clones can be verified by cotransformation assays using plasmids encoding respective bait and prey proteins. Other preferred methods for studying protein-protein interactions according to the present invention are colocalization, coimmunoprecipitation, screening of protein or (poly)peptide arrays, library screens, in vivo and in vitro binding experiments using different tags such as HIS6, TAP or FLAG.

In a preferred embodiment of the present invention's method for generating a network of direct and indirect interaction partners of a disease related protein or (poly)peptide, plasmids encoding bait proteins are transformed into a strain such as L40ccua, tested for the absence of reporter gene activity and co-transformed with a human fetal brain cDNA library. Independent transformants are plated onto minimal medium lacking tryptophan, leucine, histidine and uracil (SDIV medium) and incubated at about 30°C for 5 to 10 days. Clones are transferred into microtitre plates, optionally using a picking robot, and grown over night in liquid minimal medium lacking tryptophan and leucine (SDII medium). Subsequently, the clones are spotted onto nylon or nitrocellulose membranes placed on SDIV medium plates. After incubation for about 4 days membranes are subjected to a  $\beta$ -galactosidase ( $\beta$ -GAL) assay. Plasmids are prepared from positive clones and characterised, for example by restriction analyses and sequencing. For retransformation assays plasmids encoding bait and prey proteins are cotransformed in the yeast strain L40ccua and plated onto SDIV medium.

The term "generating a protein-protein interaction (PPI) network" means listing the interactions of all proteins interacting or binding directly or indirectly interacting the disease related (poly)peptide or protein. Preferably, this can be done by displaying the information in a matrix or a network representation. In a more preferred embodiment of the present invention's method, the protein-protein interaction network is generated by using Pivot 1.0 (Prof. Ron Shamir, Prof. Yossi Shilo, Nir Orlev; Tel Aviv University (TAU); Dep. of computer science; Ramat Aviv; Tel Aviv 69978; Israel).

In a preferred embodiment of the invention, interactions are detected by using the yeast two-hybrid system, MALDI-TOF MS or electro spray MS. Preferably, yeast strains such as strains L40ccua and L40cca, are transformed with an expression selected from the group consisting of pBTM116, pBTM117, pBTM117c, pACT2, pAS2-1, pGAD10, pGAD424, pGAD425, pGAD426, pGAD427, pGAD428.

In another preferred embodiment of the present invention's method for generating a network of direct and indirect interaction partners of a disease-related polypeptide, the method contains after step (d) the additional steps of isolating a nucleic acid molecule with homology to said nucleic acid molecule expressing the encoded protein and testing it for its activity as a modulator of huntingtin, wherein said nucleic acid molecule is DNA, RNA, cDNA, or genomic DNA. Said testing can be done in several different assays. Preferably, the testing is performed in a co-immunoprecipitation assay or an affinity chromatography-based technique. Generally, co-immunoprecipitation is performed by purifying an interacting protein complex with a single antibody specific for one protein in the protein complex and by detecting the proteins in the protein complex. The step of detection can involve the use of additional antibodies directed against proteins suspected of being trapped in the purified protein complex. Alternatively, at least one protein in the protein complex is fused to a tag sequence with affinity to a compound fixed to a solid matrix. By contacting the solid matrix with said tagged protein, further proteins binding to said protein can be purified and binding can be detected. GST or HA are preferred tags in accordance with the present invention.

In a preferred embodiment of the present invention's method, said contacting step (e) is effected in an interaction mating two hybrid approach.

In another preferred embodiment of the present invention's method, said method comprises after step (d) and before step (e) the steps of: (d') contacting (poly)peptides detected in step (d) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (d) under conditions that allow the interaction between interaction partners to occur; and (d'') detecting proteins that interact with said (poly)peptides detected in step (d').

This preferred embodiment of the invention, an additional step of identifying further interaction partners is carried out prior to the contacting of all "baits" and "preys" in one pool (step (e)). Optionally, further steps of selecting interaction partners in analogy to steps (d') and (d'') may be infected prior to the pooling/interaction step.

Diseases of particular interest for which interrelationships of disease-related proteins may be analyzed in accordance with the invention are provided in Table 5.

In yet another preferred embodiment of the present invention's method, said disease related protein is a protein suspected of being a causative agent of a hereditary (see Table 5), such as a monogenic disease.

In another preferred embodiment of the present invention's method, said disease related protein is huntingtin and said interaction partners are the interaction partners as shown in table 1 and/or table 2.

In another preferred embodiment of the present invention's method, said method comprises the step of determining the nucleotide sequence of a nucleic acid molecule encoding a direct or indirect interaction partner of the disease related protein.

In another preferred embodiment of the present invention's method, said selections of proteins are translated from a nucleic acid library.

In another preferred embodiment of the present invention's method, said selection of proteins in step (a) and/or (c) and/or (d') and/or (e) is the same selection or a selection from the same source. In another preferred embodiment of the present

invention's method, said selection of proteins in step (a) and/or (c) and/or (d') and/or (e) is a different selection or a selection from a different source.

Preferably, said source is selected from nerve cells, brain tissue, human adrenal gland, human bladder, human bone, human brain, human colon, human dorsal root ganglion, human heart, human HeLa cells, human kidney, human liver, human lung, human mammary gland, human ovary, human pancreas, human placenta, human prostate, human retina, human salivary gland, human skeletal muscle, human small intestine, human smooth muscle, human spinal cord, human spleen, human stomach, human testis, human thymus, human thyroid, human tonsil, human trachea, human uterus, human cell line HEP G2, human cell line MDA 435, human fetal brain, human fetal heart, human fetal kidney, human fetal liver, human fetal spleen, human fetal thymus, human breast tumor, human cervix tumor, human colon tumor, human kidney tumor, human lung tumor, human ovary tumor, human stomach tumor, human brain tumor and/or human uterus tumor.

In another preferred embodiment of the present invention's method, said method is performed by contacting the proteins on an array. Preferably, said array is an array allowing to detect protein-protein interaction by the principle of a biacore detector.

In another preferred embodiment of the present invention's method, said interactions are detected by using the yeast two-hybrid system. Preferably, said interactions detected by using MALDI-TOF, MS, electro spray MS or biacore.

In another preferred embodiment of the present invention's method, said method contains after step of (b), (d), (d'') or (f) the additional steps of isolating a nucleic acid molecule with homology to said cDNA expressing the encoded protein and testing it for its activity as a modulator of huntingtin, wherein said nucleic acid molecule is DNA, or RNA, and preferably cDNA, or genomic or synthetic DNA, or mRNA.

The present invention also relates to a nucleic acid molecule encoding a modulator of huntingtin, wherein said modulator is a protein selected from table 3. Figure 6 provides the amino acid sequences of the new proteins or (poly)peptides listed in table 3. The term "modulator protein of huntingtin" comprises two types of proteins within the network of proteins interacting with huntingtin. Direct interaction or binding partners of huntingtin are those proteins in the PPI network of huntingtin that directly

interact with or bind to huntingtin (see figure 2). Examples of these proteins are IKAP, HYPA, CA150, HIP1, HIP11, HIP13, HIP15, CGI-125, PFN2, HP28, DRP-1, SH3GL3, HZFH, HIP5, PIAS $\gamma$ , HIP16, GIT1, Ku70 and FEZ1. Table 2 and figure 6 provides a reference allowing to identify these proteins. The second class of proteins are indirect interaction or binding partners of huntingtin, i.e. those proteins in the PPI network of huntingtin that do not directly interact with or bind to huntingtin. Such proteins require a mediator, i.e. a direct binding partner of huntingtin to exert their huntingtin modulating function. Examples of these proteins are BARD1 or VIM, which bind to direct interaction partners of huntingtin. However, complexes of huntingtin and a direct interaction or binding partner are likely to interact with additional indirect interaction or binding partners. To summarize the above, modulator proteins of huntingtin can exert their function by direct or indirect contact to huntingtin.

The term "modulator protein", as used in the present invention, refers to a protein capable of modulating the function or physical state of a second protein and comprises proteins that enhance or reduce (inhibit) the function or activity of huntingtin. Preferably, the modulator protein is a protein having an activity selected from the group consisting of oxidoreductase activity (acting on the CH-OH group of donors, acting on the aldehyde or oxo group of donors, acting on the CH-CH group of donors, acting on the CH-NH(2) group of donors, acting on the CH-NH group of donors, acting on NADH or NADPH, acting on other nitrogenous compounds as donors, acting on a sulfur group of donors, acting on a heme group of donors, acting on diphenols and related substances as donors, acting on a peroxide as acceptor, acting on hydrogen as donor, acting on single donors with incorporation of molecular oxygen, acting on the CH-OH group of donors, acting on superoxide as acceptor, oxidizing metal ions, acting on -CH(2) groups, acting on iron-sulfur proteins as donors, acting on reduced flavodoxin as donor, acting on phosphorus or arsenic in donors, acting on x-H and y-H to form an x-y bond, other oxidoreductases), transferase activity (transferring one-carbon groups, transferring aldehyde or ketone residues, acyltransferases, glycosyltransferases, transferring alkyl or aryl groups, other than methyl groups, transferring nitrogenous groups, transferring phosphorous-containing groups, transferring sulfur-containing groups, transferring selenium-containing groups), hydrolase activity (glycosylase activity, acting on ether bonds, acting on peptide bonds, acting on carbon-nitrogen bonds (other than peptide

bonds), acting on acid anhydrides, acting on carbon-carbon bonds, acting on halide bonds, acting on phosphorus-nitrogen bonds, acting on sulfur-nitrogen bonds, acting on carbon-phosphorus bonds, acting on sulfur-nitrogen bonds, acting on carbon-phosphorus bonds; acting on sulfur-sulfur bonds, acting on carbon-sulfur bonds, lyases (carbon-carbon lyases, carbon-oxygen lyases, carbon-nitrogen lyases, carbon-sulfur lyases, carbon-halide lyases, phosphorus-oxygen lyases, other lyases), isomerases (racemases and epimerases, cis-trans-isomerases, intramolecular oxidoreductases, intramolecular transferases, intramolecular lyases, other isomerases), ligases activity (forming carbon-oxygen bonds, forming carbon-sulfur bonds, forming carbon-nitrogen bonds, forming carbon-carbon bonds, forming phosphoric ester bonds), transcription factor activity, filament protein, membrane protein and structural protein.

In a preferred embodiment, the present invention's nucleic acid molecule is DNA, or RNA, and preferably cDNA, or genomic DNA or synthetic DNA or mRNA

In another preferred embodiment of the invention, the nucleic acid molecule is double stranded or single stranded.

In another preferred embodiment of the invention, the nucleic acid molecule is of vertebrate, nematode, insect, bacterium or yeast. Preferably, the nematode is *Caenorhabditis elegans*. In another more preferred embodiment of the present invention, the insect is *drosophila*, preferably *drosophila melanogaster*. In another more preferred embodiment of the present invention, the vertebrate is human, mouse rat, *Xenopus laevis*, zebrafish.

In yet another preferred embodiment of the present invention, the nucleic acid molecule is fused to a heterologous nucleic acid molecule. In a further preferred embodiment of the present invention, the heterologous (poly)peptide encoded by said heterologous nucleic acid molecule is an immunoglobulin Fc domain.

In another preferred embodiment of the present invention the nucleic acid molecule is labeled. Labeled nucleic acid molecules may be useful for purification or detection. Suitable labels include fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-

FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine(ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g.  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ; etc. The label may also be a two stage system, where the DNA is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. In the case of amplification the label may be conjugated to one or both of the primers. The pool of nucleotides used in the amplification may also be labeled, so as to incorporate the label into the amplification product. Alternatively, the double strand formed after hybridization can be detected by anti-double strand DNA specific antibodies or aptamers etc.

In a more preferred embodiment said heterologous nucleic acid molecule encodes a heterologous polypeptide. Preferably said heterologous (poly)peptide, fused to the (poly)peptide encoded by the nucleic acid molecule of the present invention, is a DNA binding protein selected from the group consisting of GAL4 (DBP) and LexA (DBP). Also preferred in accordance with the present invention are activation domains selected from the group consisting of GAL4(AD) and VP16(AD). Also preferred are (poly)peptides selected from the group consisting of GST, His Tag, Flag Tag, Tap Tag, HA Tag and Protein A Tag.

Thus, the sequence encoding the (poly)peptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused (poly)peptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell* 37: 767 (1984).

The (poly)peptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous

functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the (poly)peptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the (poly)peptide to facilitate purification. Such regions may be removed prior to final preparation of the (poly)peptide. The addition of peptide moieties to (poly)peptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to stabilize and purify proteins.

The present invention also relates to a method of producing a vector comprising the nucleic acid molecule the present invention. Furthermore, the present invention relates to a vector produced said method.

The present invention also relates to a vector comprising the nucleic acid molecule of the present invention. Preferably said vector is a transfer or expression vector selected from the group consisting of pACT2; pAS2-1; pBTM116; pBTM117; pcDNA3.1; pcDNA1; pECFP; pECFP-C1; pECFP-N1; pECFP-N2; pECFP-N3; pEYFP-C1; pFLAG-CMV-5 a, b, c; pGAD10; pGAD424; pGAD425; pGAD427; pGAD428; pGBT9; pGEX-3X1; pGEX-5X1; pGEX-6P1; pGFP; pQE30; pQE30N; pQE30-NST; pQE31; pQE31N; pQE32; pQE32N; pQE60; pSE111; pSG5; pTET-CMV-AS; pTET-CMV-F°-AS; pTET-CMV-F°-S; pTET-CMV-MCS; pTET-CMV-S; pTK-Hyg; pTL1; pTL10; pTL-HA0; pTL-HA1; pTL-HA2; pTL-HA3; pBTM118c; pGEX-6P3; pACGHLT-C; pACGHLT-A; pACGHLT-B; pUP; pcDNA3.1-V5His; pMalc2x. Said expression vectors may particularly be plasmids, cosmids, viruses or bacteriophages used conventionally in genetic engineering plasmids, cosmids, viruses and bacteriophages used conventionally in genetic engineering that comprise the aforementioned nucleic acid. Preferably, said vector is a gene transfer or targeting vector. Expression vectors derived from viruses such as retroviruses, vaccinia virus, adeno-associated virus, herpes viruses, or bovine papilloma virus, may be used for delivery of the nucleic acid into targeted cell population. Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors; see, for example, the techniques described in Sambrook et al., Molecular Cloning A

Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel et al., Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989).

In yet a further preferred embodiment of the invention the vector contains an additional expression cassette for a reporter protein, selected from the group consisting of  $\beta$ -galactosidase, luciferase, green fluorescent protein and variants thereof.

Preferably, said vector comprises regulatory elements for expression of said nucleic acid molecule. Consequently, the nucleic acid of the invention may be operatively linked to expression control sequences allowing expression in eukaryotic cells. Expression of said nucleic acid molecule comprises transcription of the sequence nucleic acid molecule into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably mammalian cells, are well known to those skilled in the art. They usually comprise regulatory sequences ensuring initiation of transcription and, optionally, a poly-A signal ensuring termination of transcription and stabilization of the transcript, and/or an intron further enhancing expression of said nucleic acid. Additional regulatory elements may include transcriptional as well as translational enhancers, and/or naturally-associated or heterologous promoter regions. Possible regulatory elements permitting expression in eukaryotic host cells are the AOX1 or GAL1 promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. Beside elements which are responsible for the initiation of transcription such regulatory elements may also comprise transcription termination signals, such as the SV40-poly-A site or the tk-poly-A site, downstream of the nucleic acid molecule. Furthermore, depending on the expression system used leader sequences capable of directing the (poly)peptide to a cellular compartment or secreting it into the medium may be added to the coding sequence of the aforementioned nucleic acid and are well known in the art. The leader sequence(s) is (are) assembled in appropriate phase with translation, initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein, or a portion thereof, into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein

including an C- or N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDVI (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3, the Echo<sup>TM</sup> Cloning System (Invitrogen), pSPORT1 (GIBCO BRL) or pRevTet-On/pRevTet-Off or pCI (Promega).

The present invention also relates to a method of producing a host cell comprising genetically engineering cells with the nucleic acid molecule or the vector of the present invention. The present invention also relates to a host cell produced said method. Furthermore, the present invention relates to a host cell comprising the vector of the present invention. Preferably, said host cell contains an endogenous nucleic acid molecule which is operably associated with a heterologous regulatory control sequence, including the regulatory elements contained in the vector of the present invention.

The present invention also relates to a method of producing a (poly)peptide, comprising culturing the host cell of the present invention under conditions such that the (poly)peptide encoded by said polynucleotide is expressed and recovering said (poly)peptide.

The present invention also relates to a (poly)peptide comprising an amino acid sequence encoded by a nucleic acid molecule the present invention, or which is chemically synthesized, or is obtainable from the host cell of the present invention, or which is obtainable by a method the present invention.

In another preferred embodiment of the invention, the (poly)peptide or protein is of vertebrate, nematode, insect, bacterium or yeast. Preferably, the nematode is *Caenorhabditis elegans*. In another more preferred embodiment of the present invention, the insect is *drosophila*, preferably *drosophila melanogaster*. In another more preferred embodiment of the present invention, the vertebrate is human, mouse rat, *Xenopus laevis*, zebrafish.

In another preferred embodiment, the (poly)peptide of the present invention is fused to a heterologous (poly)peptide. Such a fusion protein may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of

additional amino acids, particularly charged amino acids, may be added to the N-terminus of the (poly)peptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the (poly)peptide to facilitate purification. Such regions may be removed prior to final preparation of the (poly)peptide. The addition of peptide moieties to (poly)peptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to stabilize and purify proteins.

In a preferred embodiment of the present invention, the (poly)peptide of the present invention is fused to a heterologous (poly)peptide which is an immunoglobulin Fc domain or Protein A domain. In another preferred embodiment of the present invention, the (poly)peptide the (poly)peptide is labelled. Preferably, the label is selected from the group consisting of fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine(ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g.  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ; etc. The label may also be a two stage system, where the protein or (poly)peptide is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. In another preferred embodiment of the present invention the label is a toxin, radioisotope, or fluorescent label.

In another preferred embodiment of the present invention, the (poly)peptide contains or lacks an N-terminal methionine. It is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

The present invention also relates to a protein complex comprising at least two proteins, wherein the two proteins are selected from the group of interaction partners listed in table 4. The term "protein complex" refers to a compound stably comprising at least two proteins wherein said stability allows to purify said protein complex.

In a preferred embodiment of the present invention, the protein complex comprises GIT1 and huntingtin.

The present invention also relates to an antibody specifically recognizing the (poly)peptide of the present invention or specifically reacting with the protein complex of the present invention. This antibody is characterized in not recognizing the individual components of the protein complex but rather the complex itself. As such, said antibody recognizes a combined epitope, composed of amino acids of two different proteins within the protein complex. Dissociation of the complex will be detrimental to antibody recognition. Therefore, antibody binding depends on the integrity of the protein complex. In a preferred embodiment of the present invention, the antibody is specific for a protein complex comprising GIT1 and huntingtin.

In a preferred embodiment, the antibody of the present invention is polyclonal, monoclonal, chimeric, single chain, single chain Fv, human antibody, humanized antibody, or Fab fragment

In a more preferred embodiment of the present invention the antibody is labeled. Preferably, the label is selected from the group consisting of fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine(ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g.  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ; etc. The label may also be a two stage system, where the antibody is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. In another preferred embodiment of the present invention the label is a toxin, radioisotope, or fluorescent label.

In a preferred embodiment of the present invention, the antibody is immobilized to a solid support. Preferably, the solid support may be the surface of a cell, a microtiter plate, beads or the surface of a sensor capable of detecting binding of the antibody or to the antibody.

The present invention also relates to a method of identifying whether a protein promotes huntingtin aggregation, comprising (a) transfecting a first cell with a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates; (b) co-transfected a second cell with (i) a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates; and (ii) a nucleic acid molecule encoding a candidate modulator protein identified by the methods of the present invention or a nucleic acid molecule encoding a modulator protein selected from table 1 or table 2 (c) expressing the proteins encoded by the transfected nucleic acid molecule of (a) and (b); (d) isolating insoluble aggregates of huntingtin from the transfected cell of (a) and (b); and (e) determining the amount of insoluble huntingtin aggregates from the transfected cell of (a) and (b), wherein an increased amount of huntingtin aggregates isolated from the transfected cells of (b) in comparison with the amount of huntingtin aggregates isolated from the transfected cells of (a) is indicative of a protein's activity as an enhancer of huntingtin aggregation. Preferably, the huntingtin protein or protein fragment of step (a) is HD169Q68 or HD510Q68.

The present invention also relates to a method of identifying whether a protein inhibits huntingtin aggregation, comprising (a) transfecting a first cell with a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates; (b) co-transfected a second cell with (i) a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates; and (ii) a nucleic acid molecule encoding a candidate modulator protein identified by the methods of the present invention or a nucleic acid molecule encoding a modulator protein selected from table 1 or table 2 (c) expressing the proteins encoded by the transfected nucleic acid molecule of (a) and (b); (d) isolating insoluble aggregates of huntingtin from the transfected cell of (a) and (b); and (e) determining the amount of insoluble huntingtin aggregates from the transfected cell of (a) and (b), wherein a reduced amount of

huntingtin aggregates isolated from the transfected cells of (b) in comparison with the amount of huntingtin aggregates isolated from the transfected cells of (a) is indicative of a protein's activity as an inhibitor of huntingtin aggregation. Preferably, the huntingtin protein or protein fragment of step (a) is HD169Q68 or HD510Q68 or HdexQ51.

The term "promotes" means increasing the amount of huntingtin aggregation.

Preferably said huntingtin protein or the fragments thereof is selected from the proteins listed in table 1 and/or 2. Preferably said insoluble aggregates are isolated by using a filter retardation method comprising lysing cells and boiling in 2%SDS for 5min in the presence of 100mM DDT followed by a filtration step. The presence of aggregates is detected by using specific antibodies.

In a preferred embodiment of the present invention, determining the amount of insoluble huntingtin is performed by using light scattering or size exclusion chromatography. In another preferred embodiment of the present invention prior to step (d) the cells are treated with an ionic detergent. In yet another preferred embodiment of the methods of the present invention, the huntingtin aggregates are filtered onto a membrane.

The present invention also relates to a method for identifying compounds affecting, e.g. interfering or enhancing the interaction of huntingtin or of a direct or indirect interaction partner of huntingtin comprising (a) contacting interacting proteins selected from the group of interacting proteins listed in table 1 in the presence or absence of a potential modulator of interaction; and (b) identifying compounds capable of modulating said interaction. The contacting is performed under conditions that permit the interaction of the two proteins. Sometimes more than two interacting proteins might be present in a single reaction as additional interaction partners of those listed under table 1, can be tested. However, the compound may also be a small molecule. Preferably said compounds are antibodies directed to huntingtin or to said interaction partner listed in table 1, wherein these antibodies are capable of interfering with the interaction with huntingtin. Alternatively, said compound is a peptide fragment of 10 to 25 amino acid residues of an interaction partner listed in table 1, wherein said peptide fragment is capable of interfering with the interaction

with huntingtin. In a more preferred embodiment of the present invention, said antibody is an antibody directed to GIT1. In another more preferred embodiment of the invention, said peptide fragment is a peptide fragment of GIT1 of 10 to 25 capable of interfering with the interaction of GIT1 with huntingtin. Said interfering peptide may contain additional modifications in order to increase cellular uptake, solubility or to increase stability. Such modifications are known to the person skilled in the art and need not be listed here in detail. In a preferred embodiment of the present invention, the methods for identifying a compound further comprise the steps of modeling said compound by peptidomantics and chemically synthesizing the modeled compound.

In another preferred embodiment of the present invention, the methods for identifying a compound further comprise producing said compound. In yet another preferred embodiment of the present invention, the method for identifying said compound further comprise modifying to achieve (i) modified site of action, spectrum of activity, organ specificity, and/or (ii) improved potency, and/or (iii) decreased toxicity (improved therapeutic index), and/or (iv) decreased side effects, and/or (v) modified onset of therapeutic action, duration of effect, and/or (vi) modified pharmokinetic parameters (resorption, distribution, metabolism and excretion), and/or (vii) modified physico-chemical parameters (solubility, hygroscopicity, color, taste, odor, stability, state), and/or (viii) improved general specificity, organ/tissue specificity, and/or (ix) optimized application form and route by (i) esterification of carboxyl groups, or (ii) esterification of hydroxyl groups with carbon acids, or (iii) esterification of hydroxyl groups to, e.g. phosphates, pyrophosphates or sulfates or hemi succinates, or (iv) formation of pharmaceutically acceptable salts, or (v) formation of pharmaceutically acceptable complexes, or (vi) synthesis of pharmacologically active polymers, or (vii) introduction of hydrophilic moieties, or (viii) introduction/exchange of substituents on aromates or side chains, change of substituent pattern, or (ix) modification by introduction of isosteric or bioisosteric moieties, or (x) synthesis of homologous compounds, or (xi) introduction of branched side chains, or (xii) conversion of alkyl substituents to cyclic analogues, or (xiii) derivatisation of hydroxyl group to ketales, acetals, or (xiv) N-acetylation to amides, phenylcarbamates, or (xv) synthesis of Mannich bases, imines, or transformation of

ketones or aldehydes to Schiff's bases, oximes, acetals, ketales, enoesters, oxazolidines, thiazolidines or combinations thereof.

The present invention also relates to a method of diagnosing Huntington's disease in a biological sample comprising the steps of (a) contacting the sample with an antibody specific for a protein of table 1 or 2 or an antibody specific for the protein complex of the present invention; and (b) detecting binding of the antibody to a protein complex, wherein the detection of binding is indicative of Huntington's disease or of a predisposition to develop Huntington's disease. Preferably, binding is detected by measuring the presence of a fluorescent label bound to the protein complex.

In a preferred embodiment of the present invention's method protein complex contains (a) GIT1 or (b) said antibody is specific for a protein complex containing GIT1.

The present invention also relates to a diagnostic agent/composition comprising the nucleic acid molecule of the present invention, the (poly)peptide of the present invention including/or the (poly)peptide mentioned in table 1 or 2, the antibody of the present invention, an antibody specifically reacting with a protein selected from table 2 and/or a protein selected from table 2.

Moreover, the present invention also relates to a pharmaceutical composition comprising the nucleic acid molecule of the present invention, the (poly)peptide of the present invention, the interfering compound identified with a method of the present invention, the antibody of the present invention, an antibody specifically reacting with a protein selected from table 2 and/or a protein selected from table 2.

The pharmaceutical composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient, the site of delivery of the pharmaceutical composition, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" of the pharmaceutical composition for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of pharmaceutical composition administered parenterally per dose will be in the range of about 1 µg protein /kg/day to 10 mg protein /kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg protein /kg/day, and most preferably for humans between about 0.01 and 1 mg protein /kg/day for the peptide. If given continuously, the pharmaceutical composition is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions of the invention may be administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, drops or transdermal patch), buccally, or as an oral or nasal spray. By "pharmaceutically acceptable carrier" is meant a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The pharmaceutical composition is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al., *Id.*) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release pharmaceutical composition also include liposomally entrapped protein, antibody, (poly)peptide, peptide or nucleic acid. Liposomes containing the pharmaceutical composition are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad.*

Sci. (USA) 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. (USA) 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal therapy.

For parenteral administration, in one embodiment, the pharmaceutical composition is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to (poly)peptides.

Generally, the formulations are prepared by contacting the components of the pharmaceutical composition uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes. The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) (poly)peptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG. The

proteinaceous components of the pharmaceutical composition are typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation protein or (poly)peptide salts.

The components of the pharmaceutical composition to be used for therapeutic administration must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic components of the pharmaceutical composition (poly)peptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

The components of the pharmaceutical composition ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous protein solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized protein using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical/diagnostic pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical/diagnostic compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the (poly)peptides of the components of the pharmaceutical composition invention may be employed in conjunction with other therapeutic compounds.

Finally, the present invention relates to the use of the nucleic acid molecule of the present invention, the interfering compound identified with a method of the present invention, the (poly)peptide of the present invention including/or the (poly)peptide mentioned in table 1 or 2, the antibody of the present invention, an antibody specifically reacting with a protein selected from table 2 and/or a protein selected

from table 2 for the preparation of a pharmaceutical composition for the treatment of Huntington's disease.

Tables:

Table 1:

PROTEIN-PROTEIN INTERACTIONS IN THE PPI OF HUNTINGTIN	
Baits (DBD)	Preys (AD)
BARD1	PLIP
EF1G	EF1G
HD1.7	CA150
HD1.7	HIP1
HD1.7	HYPA
HD1.7	SH3GL3
HDexQ20	CA150
HDexQ20	HYPA
HDexQ20	SH3GL3
HDexQ51	CA150
HDexQ51	HYPA
HDexQ51	SH3GL3
mp53	p53
mp53	PIASy
PIASy	SUMO-2
PIASy	SUMO-3
VIM	NEFL
VIM	VIMc
BARD1	BAIP1
BARD1	BAIP2
BARD1	BAIP3
BARD1	FEZ1
BARD1	GIT1
BARD1	HBO1
BARD1	HIP5
BARD1	HZFH
BARD1	IKAP
BARD1	mHAP1
BARD1	NAG4
BARD1	PIASy
BARD1	PTN
BARD1	SETBD1

BARD1	ZHX1
CLH-17	Ku70
CLK1	PIASy
GADD45G	BAIP3
GADD45G	CGI-125
GADD45G	CGI-74
GADD45G	EF1A
GADD45G	EF1G
GADD45G	G45IP1
GADD45G	G45IP2
GADD45G	G45IP3
GADD45G	HIP16
GADD45G	HIP5
GADD45G	LUC7B1
GADD45G	PIASy
GADD45G	PLIP
GADD45G	PTN
GADD45G	PTPK
hADA3	BAIP1
hADA3	Ku70
hADA3	MAGEH1
hADA3	PIASy
HD1.7	CGI-125
HD1.7	DRP-1
HD1.7	FEZ1
HD1.7	GIT1
HD1.7	HIP11
HD1.7	HIP13
HD1.7	HIP15
HD1.7	HIP16
HD1.7	HIP5
HD1.7	HZFH
HD1.7	IKAP
HD1.7	Ku70
HD1.7	PIASy
HDd1.0	FEZ1
HDd1.0	GIT1
HDd1.0	IKAP
HDd1.3	HZFH

HDd1.3	IKAP
HDd1.3	Ku70
HDd1.3	PIASy
HDexQ20	CGI-125
HDexQ20	HIP13
HDexQ20	HP28
HDexQ20	PFN2
HDexQ51	CGI-125
HDexQ51	HIP13
HDexQ51	HIP15
HDexQ51	HP28
HDexQ51	PFN2
HIP2	PIASy
HIP5	APP1
HIP5	BAIP1
HIP5	BAIP2
HIP5	CGI-74
HIP5	FEZ1
HIP5	GIT1
HIP5	HBO1
HIP5	HMP
HIP5	KPNA2
HIP5	mHAP1
HIP5	NAG4
HIP5	PLIP
IMPD2	PIASy
KPNB1	PIASy
KPNB1	PTN
mp53	HZFH
mp53	ZHX1
PIASy	MAPIc3
TAL1	ZHX1
TCP1G	Ku70
VIM	ALEX2
VIM	BAIP1
VIM	DRP-1
VIM	G45IP1
VIM	HBO1
VIM	HSPC232

VIM	HZFH
VIM	PIASy
VIM	SETBD1
VIM	SH3GL3
ZNF33B	mHAP1
ZNF33B	ZHX1

**Table 2 Classification of proteins in Huntington's disease interaction network**

ID	NAME	FUSION	ACCESSION	IDEN	aa MATCH	LOC
<b>Huntingtin fragments</b>						
HD1.7	huntingtin	DBD	P42858	100	1-506	N, C
HDd1.0	huntingtin	DBD	P42858	100	1-320	N, C
HDd1.3	huntingtin	DBD	P42858	100	166-506	N, C
HdexQ20	huntingtin	DBD	P42858	98	1-90	N, C
HdexQ51	huntingtin	DBD	P42858	75	1-82	N, C
<b>Transcriptional control and DNA maintenance</b>						
BARD1	BRCA1 associated ring domain protein 1	DBD	Q99728	99	1-379	N
CA150	putative transcription factor CA150	AD	O14776	93	299-629	N
GADD45G	growth arrest and DNA damage inducible protein GADD45 gamma	DBD	O95257	100	18-159	N
hADA3	ADA3 like protein	DBD	O75528	100	235-432	N
HBO1	histone acetyltransferase binding to ORC	AD	O95251	100	1-611	N
PIASy	protein inhibitor of activated STAT protein gamma (PIASy)	AD, DBD	Q8N2W9	100	5-510	N, C
HYPA	huntingtin interacting protein HYPA/FBP11 (fragment)	AD	O75400	100	8-422	C, N
HZFH	zinc finger helicase HZFH	AD, DBD	Q9Y4I0	100	1830-2000	N
IKAP	IKK complex associated protein	AD	O95163	100	1207-1332	N, C
Ku70	ATP dependent DNA helicase II, 70 kDa subunit	AD	P12956	100	298-608	N
NAG4	bromodomain containing protein NAG4	AD	Q9NP11	100	94-651	N
p53	cellular tumor antigen p53	AD	P04637	100	1-393	N
p53c	cellular tumor antigen p53 (C-terminus)	AD	P04637	100	248-393	N
mp53	cellular tumor antigen p53 (mouse)	DBD	P02340	100	73-390	N
PLIP	cPLA2 interacting protein	AD	O95624	100	5-461	N, PN
SETDB1	histone-lysine N-methyltransferase, H3 lysine-9 specific 4	AD	Q15047	100	1023-1291	N
SUMO-2	ubiquitin like protein SMT3A (SUMO-2)	AD	P55854	100	1-103	C, N
SUMO-3	ubiquitin like protein SMT3B (SUMO-3)	AD	P55855	100	1-95	C, N
ZHX1	zinc finger homeobox protein ZHX1	AD	Q9UKY1	100	145-873	N
ZNF33B	zinc finger protein 33b	DBD	Q8NDW3	100	527-778	N
<b>Cellular organization and protein transport</b>						
APP1	amyloid like protein 1 precursor	AD	P51693	100	243-555	PM, EC
CLH-17	clathrin heavy chain 1	DBD	Q00610	100	1-289	PM, V
HP28	axonemal dynein light chain (hp28)	AD	Q9BQZ6	100	3-258	CN
mHAP1	huntingtin associated protein 1 (mouse)	AD	O35668	100	3-471	C, EE
HIP1	huntingtin interacting protein 1	AD	O00291	100	245-631	C, GN
HMP	mitofillin	AD	Q16891	100	212-758	Mit
MAP1lc3	microtubule associated proteins 1A/1B light chain 3	AD	Q9H491	100	58-170	CN, MT
NEFL	light molecular weight neurofilament protein	AD	Q8IU72	100	1-543	CN, IF

PFN2	profilin II	AD	P35080	100	1-140	CN
PTN	pleiotrophin precursor (exon 1 included)	AD	P21246	100	1-168	PM, EC
SH3GL3	SH3 containing GRB2 like protein 3	AD	Q99963	100	3-347	V
KPNA2	karyopherin alpha-2 subunit	AD	P52292	100	141-529	C, N
KPNB1	karyopherin beta-1 subunit	DBD	Q14974	100	668-876	C, N
VIM	vimentin	DBD	P08670	100	1-466	CN, IF
VIMc	vimentin (C-terminus)	AD	P08670	100	190-466	CN, IF
<b>Cell signalling and fate</b>						
ALEX2	armadillo repeat protein ALEX2	AD	O60267	100	127-632	C, PM
CLK1	protein kinase CLK1	DBD	P49759	100	209-484	N
FEZ1	fasciculation and elongation protein zeta 1	AD	Q99689	100	131-392	C, PM
GIT1	ARF GTPase activating protein GIT1	AD	Q9Y2X7	98	249-761	PM, V
PTPK	protein-tyrosine phosphatase kappa precursor	AD	Q15262	100	1227-1439	PM, AJ
<b>Cellular metabolism</b>						
DRP-1	dihydropyrimidinase related protein 1 (C-terminus)	AD	Q14194	100	345-572	C
IMPD2	inosine-5'-monophosphate dehydrogenase 2	DBD	P12268	100	34-514	C
TAL1	transaldolase	DBD	P37837	100	3-337	C
<b>Protein synthesis and turnover</b>						
EF1A	translation elongation factor 1 alpha 1	AD	P04720	100	294-462	C, MT
EF1G	elongation factor 1 gamma	AD, DBD	P26641	100	2-437	C, MT
EF1Gc	elongation factor 1 gamma (C-terminus)	AD	P26641	100	123-437	C, MT
HIP2	ubiquitin conjugating enzyme E2-25 kDa	DBD	P27924	100	1-200	C, N
TCPG	T-complex protein 1, gamma subunit	DBD	P49368	100	252-544	C
<b>Uncharacterized proteins</b>						
BAIP1	BARD1 interacting protein 1 [similar to RIKEN cDNA 1810018M11]	AD	Q9BS30	100	1-226	UN
BAIP2	BARD1 interacting protein 2 [hypothetical protein]	AD	Q9H0I6	100	107-684	UN
BAIP3	BARD1 interacting protein 3 [hypothetical protein]	AD	Q96HT4	100	152-436	UN
CGI-74	CGI-74 protein	AD	Q9Y383	100	159-270	UN
CGI-125	CGI-125 protein	AD	Q9Y3C7	100	1-131	UN
G45IP1	GADD45G interacting protein 1 [hypothetical protein]	AD	Q9H0V7	100	1-340	UN
G45IP2	GADD45G interacting protein 2 [B2 gene partial cDNA, clone B2E]	AD	Q9NYA0	100	566-926	UN
G45IP3	GADD45G interacting protein 3 [OK/SW-CL.16]	AD	Q8NI70	100	3-134	UN
HIP5	huntingtin interacting protein 5 [hypothetical protein KIAA1377]	AD, DBD	Q9P2H0	100	445-988	N, C
HIP11	huntingtin interacting protein 11 [hypothetical protein]	AD	Q96EZ9	100	176-328	UN
HIP13	huntingtin interacting protein 13 [metastasis suppressor protein]	AD	Q96RX2	100	512-755	UN
HIP15	huntingtin interacting protein 15 [similar to KIAA0443 gene product]	AD	Q96D09	100	663-838	UN
HIP16	huntingtin interacting protein 16 [similar to KIAA0266 gene product]	AD	Q9BVJ6	100	585-771	UN
HSPC232	HSPC232	AD	Q9P0P6	92	1-319	UN
LUC7B1	putative SR protein LUC7B1 (SR+89)	AD	Q9NQ29	99	116-371	ER
MAGEH1	melanoma associated antigen H1	AD	Q9H213	100	1-219	UN

Abbreviations: aa, amino acids; IDEN, identity; LOC, localisation; AD, activation domain; DBD, DNA binding domain; AJ, adherens junctions; C, cytosol; CN, cytoskeleton; EC, extracellular space; EE, early endosomes; ER, endoplasmic reticulum; IF, intermediate filaments; GN, Golgi network; Mit, mitochondria; MT, microtubules; N, nucleus; PM, plasma membrane; PN, perinuclear; UN, unknown; V, vesicles; [ ], database annotation

**Table 3 New proteins in Huntington's disease interaction network**

ID	NAME	FUSION	ACCESSION	IDEN	aa MATCH	LOC
<b>Transcriptional control and DNA maintenance</b>						
BARD1	BRCA1 associated ring domain protein 1	DBD	Q99728	99	1-379	N
CA150	putative transcription factor CA150	AD	Q14776	93	299-629	N
<b>Cell signaling and fate</b>						
GIT1	ARF GTPase activating protein GIT1	AD	Q9Y2X7	98	249-761	PM, V
HSPC232	HSPC232	AD	Q9P0P6	92	1-319	UN
LUC7B1	putative SR protein LUC7B1 (SR+89)	AD	Q9NQ29	99	116-371	ER

Abbreviations: aa, amino acids; IDEN, Identity; LOC, localisation; AD, activation domain; DBD, DNA binding domain; AJ, adherens junctions; C, cytosol; CN, cytoskeleton; EC, extracellular space; EE, early endosomes; ER, endoplasmic reticulum; IF, intermediate filaments; GN, Golgi network; Mit, mitochondria; MT, microtubules; N, nucleus; PM, plasma membrane; PN, perinuclear; UN, unknown; V, vesicles; [ ], database annotation

**Table 4:**

New protein-protein interactions, found	
Baits (DBD)	Preys (AD)
BARD1	BAIP1
BARD1	BAIP2
BARD1	BAIP3
BARD1	FEZ1
BARD1	GIT1
BARD1	HBO1
BARD1	HIP5
BARD1	HZFH
BARD1	IKAP
BARD1	mHAP1
BARD1	NAG4
BARD1	PIASy
BARD1	PTN
BARD1	SETBD1
BARD1	ZHX1
CLH-17	Ku70
CLK1	PIASy
GADD45G	BAIP3
GADD45G	CGI-125
GADD45G	CGI-74

GADD45G	EF1A
GADD45G	EF1G
GADD45G	G45IP1
GADD45G	G45IP2
GADD45G	G45IP3
GADD45G	HIP16
GADD45G	HIP5
GADD45G	LUC7B1
GADD45G	PIASy
GADD45G	PLIP
GADD45G	PTN
GADD45G	PTPK
hADA3	BAIP1
hADA3	Ku70
hADA3	MAGEH1
hADA3	PIASy
HD1.7	CGI-125
HD1.7	DRP-1
HD1.7	FEZ1
HD1.7	GIT1
HD1.7	HIP11
HD1.7	HIP13
HD1.7	HIP15
HD1.7	HIP16
HD1.7	HIP5
HD1.7	HZFH
HD1.7	IKAP
HD1.7	Ku70
HD1.7	PIASy
HDd1.0	FEZ1
HDd1.0	GIT1
HDd1.0	IKAP
HDd1.3	HZFH
HDd1.3	IKAP
HDd1.3	Ku70
HDd1.3	PIASy
HDexQ20	CGI-125
HDexQ20	HIP13
HDexQ20	HP28

HDexQ20	PFN2
HDexQ51	CGI-125
HDexQ51	HIP13
HDexQ51	HIP15
HDexQ51	HP28
HDexQ51	PFN2
HIP2	PIASy
HIP5	APP1
HIP5	BAIP1
HIP5	BAIP2
HIP5	CGI-74
HIP5	FEZ1
HIP5	GIT1
HIP5	HBO1
HIP5	HMP
HIP5	KPNA2
HIP5	mHAP1
HIP5	NAG4
HIP5	PLIP
IMPD2	PIASy
KPNB1	PIASy
KPNB1	PTN
mp53	HZFH
mp53	ZHX1
PIASy	MAPIc3
TAL1	ZHX1
TCP1G	Ku70
VIM	ALEX2
VIM	BAIP1
VIM	DRP-1
VIM	G45IP1
VIM	HBO1
VIM	HSPC232
VIM	HZFH
VIM	PIASy
VIM	SETBD1
VIM	SH3GL3
ZNF33B	mHAP1
ZNF33B	ZHX1

**Table 5:**

- \* Aarskog syndrome
- \* Achromatopsia
- \* Acoustic neuroma
- \* Adrenal hyperplasia
- \* Adrenoleukodystrophy
- \* Agenesis of corpus callosum
- \* Aicardi syndrome
- \* Alagille syndrome
- \* Albinism
- \* Alopecia areata
- \* Alstrom syndrome
- \* Alpha-1-antitrypsin deficiency
- \* Alzheimer
- \* Ambiguous genitalia
- \* Androgen insensitivity syndrome(s)
- \* Anorchia
- \* Angelman syndrome
- \* Anophthalmia
- \* Apert syndrome
- \* Arthrogryposis
- \* Ataxia
- \* Autism
- \* Bardet-Biedl syndrome
- \* Basal cell carcinoma
- \* Batten disease
- \* Beckwith-Wiedemann syndrome
- \* Blepharophimosis
- \* Blind
- \* Branchio-Oto-Renal (BOR) syndrome
- \* Canavan
- \* Cancer: (ataxia telangiectasia, basal cell nevus, brain /spine, breast, colon / bowel, leukemia / lymphoma, lung, melanoma / skin, multiple endocrine neoplasia, oral, ovarian, prostate, retinoblastoma, testicular, von Hippel-Lindau, xeroderma pigmentosa)
- \* Cardiofaciocutaneous syndrome
- \* Celiac sprue
- \* Charcot-Marie-Tooth
- \* CHARGE association
- \* Chromosome anomalies - trisomy, deletions, inversions, duplications, translocations, 4p- (Wolf-Hirshhorn), 5 (cri-du-chat, 5p-), 6, 8p, 9 (trisomy 9, 9p-), 11 (11q, 11;22), 13 (trisomy 13, Patau), 15, 16 (mosaic), 18 (18q-, 18p-, ring 18, trisomy 18, tetrasomy 18p, Edwards), 21 (Down syndrome, trisomy 21), 22, X & Y [sex chromosome anomalies, Klinefelter (XXY, other), Turner (XO, other), fragile-X, other]
- \* Cleft lip and/or cleft palate
- \* Cockayne syndrome
- \* Coffin-Lowry syndrome
- \* Coffin-Siris syndrome
- \* Congenital heart defects

- \* Connective tissue conditions
- \* Cooley anemia
- \* Conjoined twins
- \* Cornelia de Lange syndrome
- \* Costello syndrome
- \* Craniofacial conditions
- \* Cri-du-Chat (5p-)
- \* Cystic fibrosis
- \* Cystinosis
- \* Cystinuria
- \* Dandy-Walker syndrome
- \* Deaf / hard of hearing
- \* Dermatological (skin) conditions
- \* Developmental delay / mental retardation
- \* DiGeorge syndrome
- \* Down syndrome
- \* DRPLA
- \* Dubowitz syndrome
- \* Dwarfism/ short stature
- \* Dysautonomia
- \* Dystonia
- \* Ectodermal dysplasia
- \* Ehlers Danlos syndrome
- \* Endocrine Conditions
- \* Epidermolysis bullosa
- \* Facial anomalies, disfigurement
- \* Fanconi anemia
- \* Fetal alcohol syndrome and effects
- \* FG syndrome
- \* Fragile-X syndrome
- \* Friedreich ataxia
- \* Freeman Sheldon syndrome
- \* Galactosemia
- \* Gardner syndrome
- \* Gastroenterology conditions
- \* Gaucher disease
- \* Glycogen storage disease
- \* Goldenhar syndrome
- \* Gorlin syndrome
- \* Hallerman Streiff syndrome
- \* Hearing problems
- \* Heart conditions
- \* Hemochromatosis
- \* Hemophilia
- \* Hemoglobinopathies
- \* Hereditary hemorrhagic telangiectasia
- \* Hereditary spastic paraplegia
- \* Hermansky-Pudlak syndrome
- \* Hirschsprung anomaly
- \* Holoprosencephaly

- \* Huntington disease
- \* Hydrocephalus
- \* Ichthyosis
- \* Immune deficiencies
- \* Incontinentia pigmenti
- \* Infertility
- \* Intestinal problems
- \* Joseph disease
- \* Joubert syndrome
- \* Kabuki syndrome
- \* Kidney conditions
- \* Klinefelter syndrome
- \* Klippel-Feil syndrome
- \* Klippel-Trenaunay syndrome
- \* Langer-Giedion syndrome
- \* Laurence-Moon-Biedl syndrome
- \* Leber Optic Atrophy
- \* Leigh disease
- \* Lesch-Nyhan syndrome
- \* Leukodystrophy [Adrenoleukodystrophy (ALD), Alexanders Disease, CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts & Leukoencephalopathy), Canavan Disease (Spongy Degeneration), Cerebrotendinous Xanthomatosis (CTX), Globoid Cell (Krabbes) Leukodystrophy, Metachromatic Leukodystrophy (MLD), Ovarioleukodystrophy , Pelizaeus-Merzbacher Disease, Refsum Disease, van der Knaap syndrome, Zellweger syndrome]
- \* Limb anomalies [missing arm(s) or leg(s), Poland anomaly, other]
- \* Lissencephaly [Isolated Sequence (ILS), X-Linked (XLIS), Subcortical Band Heterotopia (SBH), Miller-Dieker syndrome (MDS), Microcephaly, Microlissencephaly (MLIS), Norman-Roberts syndrome (NRS), With Cerebellar Hypoplasia (LCH), Polymicrogyria (PMG), Schizencephaly (SCH), Muscle-Eye-Brain (MEB) Disease, and Walker-Warburg syndrome (WWS), 17p13.3 deletion]
- \* Liver conditions (biliary atresia, Alagille syndrome, alpha-1 antitrypsin, tyrosinemia, neonatal hepatitis, Wilson disease)
- \* Lowe syndrome
- \* Lung / pulmonary conditions
- \* Lymphedema
- \* Maffucci syndrome(Ollier, multiple cartilaginous enchondromatosis)
- \* Malignant hyperthermia
- \* Maple syrup urine disease
- \* Marinesco-Sjogren Syndrome
- \* Marfan syndrome
- \* Menke syndrome
- \* Mental retardation / developmental delay
- \* Metabolic conditions (carbohydrate deficient glycoprotein syndrome (CDGS), diabetes insipidus, Fabry, galactosemia, glucose-6-phosphate dehydrogenase (G6PD), fatty acid oxidation disorders, glutaric aciduria, hypophosphatemia, Krabbe, lactic acidosis, lysosomal storage diseases, mannosidosis, maple syrup urine, mitochondrial, neuro-metabolic, organic acidemias, PKU, purine, pyruvate dehydrogenase deficiency, urea cycle conditions, vitamin D deficient rickets)

- \* Miscarriage, stillbirth, infant death
- \* Mitochondrial conditions (Alpers, Barth, beta-oxidation defects, carnitine deficiency, CPEO, Kearns-Sayre, lactic acidosis, Leber optic neuropathy, Leigh, LCAD, Luft, MCAD, MAD, glutaric aciduria, MERRF, MNGIE, NARP, Pearson, PHD, SCAD, NADH-CoQ reductase, succinate dehydrogenase, Complex III, Complex IV, COX, Complex V, other)
- \* Moebius syndrome
- \* Mucolipidosis, type IV (ML4)
- \* Mucopolysaccharidosis (Hunter syndrome, Hurler syndrome, Maroteaux-Lamy syndrome, Sanfilippo syndrome, Scheie syndrome, Morquio syndrome, other)
- \* Multiple hereditary exostoses
- \* Muscular dystrophy /atrophy (neuromuscular conditions including: Duchenne, facioscapulohumeral, Charcot Marie Tooth, spinal muscular atrophy, other)
- \* Myotonic dystrophy
- \* Nager & Miller syndromes
- \* Nail Patella syndrome
- \* Narcolepsy
- \* Neurologic conditions (neuro-metabolic, neurogenetics, neuromuscular, other)
- \* Neurofibromatosis (von Recklinghausen)
- \* Neuromuscular conditions
- \* Niemann-Pick disease
- \* Noonan syndrome
- \* Opitz syndromes [Opitz-Frias, Opitz FG (Opitz-Kaveggia), Opitz-C (Trigonocephaly)]
- \* Organic acidemias
- \* Osler-Weber-Rendu syndrome
- \* Osteogenesis imperfecta
- \* Oxalosis & hyperoxaluria
- \* Pallister-Hall syndrome
- \* Pallister-Killian syndrome (tetrasomy 12p, Teschler-Nicola syndrome)
- \* Parkinson's disease
- \* Periodic paralysis
- \* Phenylketonuria (PKU)
- \* Polycystic kidney disease
- \* Popliteal pterygium syndrome
- \* Porphyria
- \* Prader-Willi syndrome
- \* Progeria (Werner, Hutchinson-Gilford, Cockayne, Rothmond-Thomson syndromes)
- \* Proteus syndrome
- \* Prune belly syndrome
- \* Pseudoxanthoma elasticum (PXE)
- \* Psychiatric conditions
- \* Refsum disease
- \* Retinal degeneration
- \* Retinitis pigmentosa (retinal degenerative diseases, Usher syndrome)
- \* Retinoblastoma
- \* Rett syndrome
- \* Robinow syndrome
- \* Rubinstein-Taybi syndrome
- \* Russell-Silver syndrome

- \* SBMA
- \* SCA
- \* Schizencephaly
- \* Sex chromosome anomalies (47,XXY, 47,XXX, 45,X and variants, 47,XYY)
- \* Shwachman syndrome
- \* Sickle cell anemia
- \* Skeletal dysplasia
- \* Smith-Lemli-Opitz syndrome (RHS syndrome)
- \* Smith-Magenis syndrome (17p-)
- \* Sotos syndrome
- \* Spina bifida (myelomeningocele, neural tube defects)
- \* Spinal muscular atrophy (Werdnig-Hoffman, Kugelberg-Welander)
- \* Stickler / Marshall syndrome
- \* Sturge-Weber
- \* Tay-Sachs disease / other (dysautonomia, dystonia, Gaucher, Niemann Pick, Canavan, Bloom)
- \* Thalassemia (Cooley anemia)
- \* Thrombocytopenia absent radius syndrome
- \* Tourette syndrome
- \* Treacher Collins syndrome (craniofacial)
- \* Trisomy (21, 18, 13, 9, other, see chromosome syndromes)
- \* Tuberous sclerosis
- \* Turner syndrome
- \* Twins / triplets / multiple births
- \* Unknown disorders
- \* Urea cycle conditions
- \* Usher syndrome
- \* VATER association
- \* Velo-cardio-facial syndrome (Shprintzen, DiGeorge, 22q deletion)
- \* Visual impairment / blind
- \* Von Hippel-Lindau syndrome
- \* Waardenburg syndrome
- \* Weaver syndrome
- \* Werner syndrome
- \* Williams syndrome
- \* Wilson disease (hepatolenticular degeneration)
- \* Xeroderma pigmentosum
- \* Zellweger syndrome

The figures show:

**Figure 1** Identification of two-hybrid interactions connected to HD. a, Schematic representation of the screening strategy. b, Identification of interactions by systematic interaction mating. Upper panel: Selection of diploid yeast clones by transfer on minimal medium lacking leucine and tryptophan (SDII). Lower panel: Two-hybrid selection of interactions on minimal medium lacking leucine, tryptophan, histidine and

uracil (SDIV) after 5 days of growth at 30°C. The prey proteins HP28 (A5), SH3GL3 (A7), CA150 (B9), HIP15 (B10), PFN2 (B11), HIP13 (C1), CGI125 (C12) and HYPA (D1) were identified as HDexQ51 interactors.

**Figure 2** Protein interaction network for Huntington's disease. a, Matrix of 117 two-hybrid interactions between 21 bait and 49 prey proteins. b, Yeast two-hybrid interactions depicted as network using the software Pivot 1.0. In total, 96 interactions and 61 distinct proteins are depicted. In addition, dimers of EF1G, VIM and p53 are shown.

**Figure 3** Systematic validation of two-hybrid interactions by *in vitro* binding experiments. GST-fusion proteins (baits) immobilised on glutathione agarose beads were incubated with COS1 cell extracts containing HA-tagged prey proteins. After extensive washing of the beads, bound proteins were eluted and analysed by SDS-PAGE and immunoblotting using anti-HA antibody.

**Figure 4** Identification of network proteins stimulating htt aggregation. a, Filter retardation assay. Protein extracts were prepared from HEK293 cells coexpressing HD169Q68 and network proteins as indicated. The aggregated proteins retained on the filter were detected with anti htt antibody (CAG53b) and anti-GIT1 antibody. b, Coimmunoprecipitation of HD510Q68 and GIT1 from COS1 cell extracts. Extracts were incubated with anti-GIT1 or preimmune serum. Immunoprecipitated material was analysed by immunoblotting using htt- antibody 4C8 and anti-HA antibody. c, Coimmunoprecipitation of htt and GIT1 from human brain extracts. Protein complexes containing GIT1 were pulled-down with increasing amounts of anti htt antibodies, but not with corresponding preimmune sera. d, Analysis of subcellular localisation of HD510Q68 and GIT1 by immunofluorescence microscopy. COS1 cells were transfected with the indicated constructs and immunolabelled with 4C8 anti htt antibody coupled to Cy3-conjugated antibody (red) and with anti-HA antibody coupled to FITC-conjugated antibody (green). Nuclei were counterstained with Hoechst (blue). Colocalisation of HD510Q68 and GIT1 is illustrated by yellow colour of the insoluble aggregates. Scale bars, 10 µm.

**Figure 5** Detection of GIT1 in brains of R6/1 transgenic mice and HD patients. a, Sections of striatum and cortex of R6/1 mice brains labelled with anti-GIT1 and anti-

htt (EM48) antisera. Arrows point to nuclear inclusions. b, Inclusions in cortex of HD patients are labelled with anti-htt (2B4) and anti-GIT1 antibodies. Arrows indicate neuronal inclusions, recognized by anti-htt (2B4) and anti-GIT1 antibodies. Scale bars, 20 µm. c, Colocalisation of GIT1 and htt in the cortex of HD patients detected by immunofluorescence microscopy.

**Figure 6** Amino acid sequence of the interacting proteins of the PPI of huntingtin.

The examples illustrate the invention:

### **Examples 1: Particular methods and material used in the Examples**

#### **• Antibodies, strains and plasmids**

A polyclonal antibody (pAb) against GIT1 was generated by injection of affinity purified His<sub>6</sub>-tagged GIT1 (residues 368-587) into a rabbit. The htt-specific pAb CAG53b and HD1 were described <sup>13,14</sup>. Commercially available antibodies were anti-GST pAb (Amersham Pharmacia), anti-GIT1 pAb (Santa Cruz Biotechnology), anti-HA monoclonal antibody 12CA5 (mAb) (Roche Diagnostics), anti htt pAb EM48 <sup>47</sup>, anti htt mAb 2B4 <sup>48</sup> and anti htt mAb 4C8 (Chemicon). As secondary antibodies for immunofluorescence microscopy Cy3- and FITC-conjugated IgGs (Jackson ImmunoResearch) were used.

The yeast strains used as two-hybrid reporters were L40ccua [MAT $\alpha$  his3Δ200 trp1-901 leu2-3,112 LYS2::(lexAop)<sub>4</sub>-HIS3 ura3::(lexAop)<sub>8</sub>-lacZ ADE2::(lexAop)<sub>8</sub>- URA3 GAL4 gal80 can1 cyh2] and L40cc $\alpha$  [MAT $\alpha$  his3Δ200 trp1-910 leu2-3,112 ade2 LYS2::(lexAop)<sub>4</sub>-HIS3 URA3::(lexAop)<sub>8</sub>-lacZ GAL4 gal80 can1 cyh2]. Both strains are derivatives of L40c <sup>17</sup>. Two-hybrid vector maps are available at <http://www.mdc-berlin.de/neuroprot/labequip.htm>. Plasmids pHD510Q17 and pHD510Q68 were generated by insertion of fragments coding for HD510Q17 and HD510Q68 into pcDNA-I (Invitrogen). pHD169Q68 was derived from pHD510Q68 by deletion of the Xhol-Xhol fragment encoding aa 170-510 of human htt.

#### **• Library screening**

Plasmids encoding bait proteins were transformed into the strain L40ccua, tested for the absence of reporter gene activity and cotransformed with a human fetal brain cDNA library (Clontech). For each transformation  $1 \times 10^6$  independent transformants were plated onto minimal medium lacking tryptophan, leucine, histidine and uracil (SDIV medium) and incubated at 30°C for 5 to 10 days. Clones were picked into microtitre plates using a picking robot and grown over night in liquid minimal medium lacking tryptophan and leucine (SDII medium). Then, they were spotted onto nylon or nitrocellulose membranes placed on SDIV medium plates. After incubation for 4 days

membranes were subjected to a  $\beta$ -galactosidase ( $\beta$ -GAL) assay. Plasmids were prepared from positive clones and characterised by restriction analyses and sequencing. For retransformation assays plasmids encoding bait and prey proteins were cotransformed in the yeast strain L40ccua and plated onto SDIV medium.

- **Array mating screen**

Plasmids encoding bait and prey proteins were transformed into strains L40ccua and L40cc $\alpha$ , respectively. L40cc $\alpha$  clones were arrayed in 96-well microtitre plates and mixed with a single L40ccua clone for interaction mating. Diploid cells were transferred by a robot (Beckman, Biomek® 2000) onto YPD medium plates and, after incubation for 24 h at 30°C, onto SDII medium plates for additional 72 h at 30°C. For two-hybrid selection diploid cells were transferred onto SDIV medium plates with and without nylon or nitrocellulose membranes and incubated for 5 days at 30°C. The nylon or nitrocellulose membranes were subjected to the  $\beta$ -GAL assay. Positive clones were verified by cotransformation assays using plasmids encoding respective bait and prey proteins.

- **Protein expression and verification assays**

For verification experiments cDNA fragments encoding baits and preys were subcloned into pGEX derivatives (Stratagene) or pTL-HA<sup>18</sup>. GST fusion proteins were expressed in *E. coli* BL21-codon PlusTM RP (Stratagene) and affinity purified on glutathione agarose beads (Sigma) using standard protocols<sup>17</sup>. COS1 cells were transfected with mammalian expression plasmids and lysed as described<sup>18</sup>. For *in vitro* binding assays, 30  $\mu$ g of GST or GST fusion protein were immobilized on glutathione agarose beads and incubated with 500  $\mu$ g protein extract prepared from COS1 cells expressing a HA-tagged fusion protein for 2 h at 4°C in binding buffer [50 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol, 1 % NP-40, 1 mM EDTA, 20 mM NaF, 1 mM DTT, 0.1 % Triton X-100, protease inhibitors (Roche Diagnostics)]. After centrifugation and extensive washing of the beads bound proteins were eluted and analysed by SDS-PAGE and Western blotting.

Coimmunoprecipitation experiments were performed as described by Sittler *et al.*<sup>18</sup>. For immunofluorescence microscopy COS1 cells were grown on cover slips and

cotransfected with pcDNA-HD510Q68 and pTL-HA-GIT1. 40 h post transfection cells were fixed with 2% paraformaldehyde. Standard protocols for staining with appropriate primary and secondary antibodies were used<sup>18</sup>.

- **Filter Retardation Assay**

HEK293 cells coexpressing HD169Q68 and GIT1, PIASy, HIP5, HP28, PFN2, FEZ1 or BARD1 were harvested 48 h post transfection. Cells were lysed as described<sup>18</sup> and boiled in 2% SDS, 100 mM DTT for 5 min. Aliquots containing 50, 25 and 12.5 µg of total protein were used for filtration on a cellulose acetate membrane<sup>14</sup>. SDS-resistant aggregates were detected using anti-CAG53b or anti-GIT1 antibodies.

- **Immunocytochemistry**

Mice were deeply anaesthetised and perfused through the left cardiac ventricle with 4% paraformaldehyde in 0.1 M phosphate buffer. Brains were removed and postfixed overnight in 4% paraformaldehyde. Sections were processed for immunocytochemistry as described<sup>47</sup>. pAb EM48 (1:1000) and affinity purified anti-GIT1 pAb (1:100) were used as primary antibodies.

Six human HD and 5 control brains were used in this study. Two HD cases were classified as grade 3 and four cases as grade 4 of neuropathological severity. For immunolabelling standard protocols were used<sup>48</sup>. 2B4 mAb (1:200) and affinity purified GIT1 pAb (1:50) were used as primary antibodies.

#### **Example 2: Two-hybrid screens and data management**

To generate a PPI network for HD we used a combination of library and matrix yeast two-hybrid screens (Fig. 1a). First, 50 selected cDNAs encoding proteins potentially involved in HD including 10 different htt fragments were cloned into a DNA binding domain vector for expression of LexA fusion proteins (baits). The resulting plasmids were introduced into yeast strain L40ccua, which carries three reporter genes, *HIS3*, *URA3* and *lacZ*, for two-hybrid interaction analyses. Forty baits did not activate the reporters by themselves and were used individually for cotransformation screening of a human fetal brain cDNA library expressing GAL4 activation domain hybrids (preys). In each screen,  $1 \times 10^6$  auxotrophic transformants were tested on selective plates,

and 1-50 positive colonies were typically obtained. Restriction analyses and sequencing identified preys that together with their respective baits repeatedly activated the reporter genes. Starting with 40 baits in the first round of cotransformation screens we identified 34 PPIs for 10 baits (Table 1).

In the second round of screens, 12 cDNA fragments encoding preys identified in the first screen were subcloned into a DNA binding domain vector. The resulting baits were tested for autoactivation and 10 were screened against a human fetal brain cDNA library. Four of the 10 proteins revealed additional 13 PPIs.

Finally, an array mating screen was performed to connect all baits and preys identified in the transformation screens. For this assay, MAT $\alpha$  yeast cultures were transformed with plasmids encoding prey proteins and arrayed in 96-well microtitre plates for interaction mating with individual MAT $\alpha$  strains expressing bait proteins. Using this strategy each bait was individually tested for interaction with every prey in the array. Diploid yeast clones, formed by mating on YPD plates, were selected on agar SDII plates, and further transferred by a spotting robot on SDIV plates to select for Y2H interactions (Fig. 1b). We examined 3500 pairwise combinations of baits and preys in the mating assay and identified additional 70 PPIs. These interactions could be confirmed in cotransformation assays (Table 5).

Table 5:

**Summary of two-hybrid screens**

Screen	baits	preys	baits yielding	interactions
	screened	screened	interactions	identified
1st transformation screen	40	$4 \times 10^7$	10	34
2nd transformation screen	10	$1 \times 10^7$	4	13
Array mating screen	50	70	21	70

Thus, the combination of cDNA library and array mating screens proved powerful in establishing a highly connected PPI network linked to htt.

Sequence analysis of the cDNAs encoding bait and prey proteins revealed ORFs ranging from 82 to 728 amino acids in size (Table 2). In a systematic Blast search 60 out of the 67 proteins identified were identical to a SwissProt or TrEMBL protein entry (<http://us.expasy.org/sprot/>). The remaining 7 proteins showed 75-99 % identity to its best fit and either contained single amino acid substitutions, variable polyQ lengths or small regions of sequence variation. Uncharacterised proteins were named according to their interaction partners. Each ORF was further examined for consensus protein domains using the FprintScan, HMMPfam, HMMSmart, ProfileScan, and BlastProDom programs providing useful hints to protein function. For example, the protein BAIP1 (BARD1 interacting protein 1) possesses a Zn-finger-like PHD finger that is believed to be important for chromatin-mediated transcriptional regulation. Similarly, domain searches for BAIP2 (BARD1 interacting protein 2) revealed a BTB/POZ domain, a motif found in developmentally regulated zinc finger proteins of the Kelch family of actin-associated proteins. Thus, BAIP2 could potentially mediate the association of BARD1 with the actin cytoskeleton.

### **Example 3: Analysis and functional assignment of the two-hybrid data**

Our two-hybrid screens identified a total of 117 PPIs between 70 protein fragments. As a result of the iterative two-hybrid strategy all interactions could be depicted in a single large network. The number of interactions identified for each bait varied from 1 to 18, with each protein having 1.6 interaction partners on average. In order to display the PPI data, both matrix and network representations were used (Fig. 2). The matrix shows, in addition to the two-hybrid interactions, previously reported interactions and interactions verified by independent methods (Fig. 2a). In comparison, the network view allows to immediately recognize local PPI patterns and paths connecting two proteins in the network (Fig. 2b). Interestingly, proteins such as htt, BARD1, GADD45G, HIP5, PIASy or VIM interact with more than 11 other proteins forming nodes within the HD network, while 30 proteins have only one interaction partner and thus are located at the periphery of the network (Fig. 2b). Indeed, all other proteins are embedded in many bi-fan motifs and multiple circular interaction clusters that have been interpreted to be an indication for biological relevance<sup>11,19</sup>. Schwikowski et al.<sup>20</sup> defined network proteins, which are separated

by no more than two other proteins, as being part of a functional cluster. In this respect all proteins in our network form a functional cluster with htt.

We assigned a subcellular localisation to each protein by examining various sources of literature and based on available experimental data we grouped the proteins into six broad functional categories (Fig. 2a, Table 2).

Eighteen proteins in the HD network are involved in transcriptional regulation or DNA maintenance (Fig. 2a). The second largest group, 14 proteins, includes mainly cytoskeletal and transport proteins. We assigned 5 proteins to cellular signalling and fate, another 4 proteins to protein synthesis and turnover, and 3 proteins to cellular metabolism. Being part of 41 interactions, 16 proteins of unknown function were identified.

For the analysis of htt PPIs, as much as 40 out of 117 interactions (34,2%) included a htt fragment (Fig. 2a). In total, 19 different htt interacting partners from various functional groups were detected, 4 proteins had been previously described and 6 involved proteins of unknown function. Surprisingly, most htt partners (6) are involved in transcriptional regulation and DNA maintenance, but others function in cell organization and transport (4), cellular signalling (2), or cellular metabolism (1), suggesting that htt functions in different subcellular processes.

The current hypothesis that htt has a function in transcriptional regulation is inferred from its interactions with transcriptional activators, coactivators or repressors<sup>21</sup>. In agreement with previous reports, binding of htt to CA150<sup>22</sup> and HYPA<sup>23</sup> has been detected in our screens. In addition, new connections to nuclear proteins such as SETBD1, PLIP and HBO1 were found. These multidomain proteins act on histones and are known modulators of chromatin structure and gene expression. Similarly, the zinc finger bromo domain containing proteins BARD1, NAG4, HZFH, ZHX1, ZNF33B play a role in transcriptional control. The protein IKAP directly interacts with htt and was recently shown to be part of a complex regulating RNA polymerase II activity<sup>24</sup>. Htt also interacts with PIASy, which inhibits transcription factor STAT-mediated gene activation<sup>25</sup>. PIASy functions as SUMO E3 ligase for the Wnt-responsive transcription factor LEF1, inhibiting its activity via sumoylation<sup>26</sup>. This suggests that PIASy catalysed sumoylation of transcription factors could represent a general

mechanism in repression of gene expression. The binding of PIASy to htt indicates that htt may itself be a substrate for sumoylation. Alternatively, it could influence the sumoylation of other transcription factors. Thus, our data extend the nuclear role of htt and provide additional leads for its involvement in transcriptional regulation.

Another large group of htt interactors identified here are proteins that function in cellular organization and vesicle transport. We report a new interaction between htt and dynein light chain (HP28), a component of the dynein/dynactin motor protein complex. Interestingly, the p150<sup>Glued</sup> subunit of dynactin is linked to the htt-associated protein HAP1<sup>16,27</sup>. Our observation that htt directly binds to HP28 underscores the potential scaffolding role of htt/HAP1 in dynein/dynactin driven retrograde vesicle transport along microtubules in axons.

The htt interacting protein HIP1 anchors clathrin-coated vesicles to the cytoskeleton via its actin-binding domain, a link crucial for synaptic vesicle endocytosis<sup>28</sup>. Here, a new PPI between htt and profilin II (PFN2)<sup>29</sup> was detected. PFN2, a protein enriched in neurons, modulates actin polymerization *in vitro* and is involved in endocytosis via association with scaffolding proteins<sup>29</sup>. The htt-PFN2 connection adds support to a potential role of htt in modulation of both actin polymerization and vesicle transport processes.

Currently, for the function of 6 htt interactors, including HIP5, no genetic or biochemical evidence is available (Table 2). We found that HIP5 binds to htt as well as to karyopherin α (KPNA2). KPNA2 serves as an adapter for karyopherin β (KPNB1), which transports NLS-tagged proteins into the nucleus<sup>30</sup>. Thus, HIP5 might take this route to the nucleus. Interestingly, HEAT or armadillo (ARM) repeats, forming α-helical structures in KPNA2 and KPNB1 are also present in htt<sup>31</sup>. Therefore, the complexes between KPNA2 and HIP5 as well as between htt and HIP5 could be similar in terms of protein structure. It is tempting to further speculate that htt participates in nucleocytoplasmic transport.

### Example 3: Verification of PPIs

Comparison with literature-cited interactions revealed that more than 80% of the two-hybrid interactions identified here are novel. For all network bait and prey proteins only 24 PPIs have been reported previously using two-hybrid methods,

coimmunoprecipitations or affinity chromatography-based techniques; 18 of these were confirmed in our Y2H assays (Fig 2a, Table 2). Failure to detect interactions may result from the high stringency of our particular two-hybrid system. However, in most cases the occurrence of false negatives can be explained by the lack of essential domains in one of the protein fragments used. For example, an interaction between p53 and hADA3 has been described<sup>32</sup>, with the first 214 amino acids of hADA3 being essential for this interaction. It escaped our two-hybrid analysis, because a C-terminal hADA3 fragment (amino acids 235-432) was used. For the same reason, an interaction between p53 and BARD1 or between KPNA2 and KPNB1 was not observed.

Beside false negatives, the two-hybrid assay is also prone to create false positive results<sup>9</sup>. Addressing this issue we performed a series of pull-down and overlay assays and thereby confirmed several of the two-hybrid PPIs independently. Proteins were expressed as GST-fusions in *E. coli* and as HA-fusions in COS1 cells. After immobilization of the GST-fusion protein to beads or nitrocellulose membranes the respective partner was affinity-purified from a COS1 cell extract and binding was detected by immunoblotting. Using these assays, 22 physical interactions, central to the HD network, were verified (Fig. 2a). The results of some *in vitro* GST pull-down assays are shown in Fig. 3. For example HD510Q17 interacts with HIP1, GIT1, PIASy, FEZ1 and HIP11, and HIP5 binds to HD510Q68, GIT1, HBO1, PLIP and FEZ1 (Fig. 3). In total, 35 two-hybrid interactions were verified independently either in previous studies or by our *in vitro* binding assays (Fig. 2a).

#### **Example 4: GIT1 promotes htt aggregation *in vivo***

The formation of insoluble polyQ-containing protein aggregates is a pathological hallmark of HD. Several lines of evidence link htt aggregation to disease progression and the development of motor symptoms. We screened network proteins for their potential to enhance htt aggregation in a cell-based aggregation assay<sup>14</sup>. In this assay, formation of SDS-insoluble htt aggregates in mammalian cells, that have been cotransfected with constructs encoding an N-terminal htt fragment with 68 glutamines (HD169Q68) and a network protein of interest, is monitored by filter retardation<sup>14</sup>. HD169Q68 *per se* has only a low propensity to form insoluble aggregates in HEK293 cells. However, as shown in Fig. 4a coexpression of the htt-interacting protein GIT1

strongly promotes the formation of HD169Q68 aggregates, whereas coexpression of PIASy, HIP5, HP28, PFN2, FEZ1 and BARD1 has no discernable effect. Thus, GIT1 is a potential modifier of HD pathogenesis, which may influence the rate of formation of insoluble htt aggregates *in vivo*.

Furthermore, probing of the insoluble HD169Q68 aggregates with an anti-GIT1 antibody revealed that GIT1 does not only stimulate aggregation but is also an integral part of the insoluble aggregates (Fig. 4a). This suggests that GIT1 promotes aggregation through direct binding to mutant htt.

The interaction between GIT1 and htt was confirmed by coimmunoprecipitation from COS1 cells transfected with constructs encoding HD510Q68 and HA-GIT1. Forty hours post transfection cell extracts were prepared and treated with antiserum against GIT1. HD510Q68 and HA-GIT1 were detected in the immunoprecipitate on Western blots with anti htt antibody 4C8 and anti-HA antibody 12CA5, respectively (Fig. 4b).

The GIT1 htt interaction was also detected in human brain. Protein extracts prepared from human cortex were treated with the anti htt antibodies CAG53b and HD1, and the precipitate was probed for the presence of GIT1 (Fig. 4c). Full length GIT1, migrating at about 90 kDa<sup>33</sup>, was precipitated by both anti htt antibodies in a concentration dependent manner, indicating the existence of a complex between htt and GIT1 in neurons.

Finally, we performed colocalisation studies of htt and GIT1 in COS1 cells using immunofluorescence microscopy. In cells expressing HD510Q68 or GIT1 alone a diffuse cytoplasmic staining was observed for each protein (Fig. 4d). However, when GIT1 and mutant htt were coexpressed, large perinuclear structures, most likely reflecting protein aggregates, appeared almost exclusively. These structures contained both GIT1 and htt. The images further substantiate the findings that GIT1 and htt bind to each other and that GIT1 is a potent enhancer of mutant htt aggregation.

#### **Example 5: GIT1 localises to htt aggregates in HD transgenic mouse and patient brains**

The finding of colocalisation of htt and GIT1 within aggregates in transfected COS1 cells suggests that GIT1 might also be a component of htt aggregates *in vivo*. To investigate this possibility we first assessed the distribution of GIT1 in brains of R6/1 transgenic mice expressing a human htt exon 1 protein with 150 glutamines<sup>34</sup>. In wildtype mice, GIT1 immunoreaction product was found diffuse in the cytoplasm and nuclei of neurons throughout the brain. In R6/1 brains, in addition to the diffuse staining, GIT1 immunoreactivity was also present in large nuclear and cytoplasmic puncta similar to htt aggregates (Figure 5a). To further confirm these data, we examined the subcellular distribution of GIT1 in cortex from HD patient brains and healthy individuals (Fig. 5b). In patient brains, GIT1 antibodies labelled neuronal nuclear inclusions as well as neuropil aggregates characteristic of HD brains<sup>35</sup>. In contrast, neurons from control brains only showed a diffuse nuclear and cytoplasmic GIT1 immunostaining. In fact, in colocalisation studies performed in HD brain sections, GIT1 positive aggregates were also labelled with anti htt antibody 2B4, indicating that both proteins coaggregated *in vivo* (Fig. 5c). This observation raises the possibility that an alteration of the neuronal GIT1 subcellular distribution contributes to HD pathogenesis.

**Additional references:**

1. Tucker, C. L., Gera, J. F. & Uetz, P. Towards an understanding of complex protein networks. *Trends Cell. Biol.* **11**, 102-106 (2001).
2. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294 (1998).
3. Legrain, P., Wojcik, J. & Gauthier, J. M. Protein-protein interaction maps: a lead towards cellular functions. *Trends Genet.* **17**, 346-352 (2001).
4. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).
5. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627 (2000).
6. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569-74 (2001).
7. Rain, J. C. et al. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211-5 (2001).
8. Bader, G. D. & Hogue, C. W. Analyzing yeast protein-protein interaction data obtained from different sources. *Nature Biotechnol.* **20**, 991-997 (2002).
9. von Mering, C. et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403 (2002).
10. Walhout, A. J. et al. Integrating Interactome, Phenome, and Transcriptome Mapping Data for the *C. elegans* Germline. *Curr. Biol.* **12**, 1952-1958 (2002).
11. Walhout, A. et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116-122 (2000).
12. Fromont-Racine, M., Rain, J. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277-282 (1997).
13. Davies, S. W. et al. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537-548 (1997).
14. Scherzinger, E. et al. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90**, 549-558 (1997).
15. Cattaneo, E. et al. Loss of normal huntingtin function: new developments in Huntington's disease research. *Trends Neurosci.* **24**, 182-188 (2001).
16. Li, X. et al. A huntingtin-associated protein enriched in brain with implications for pathology. *Nature* **378**, 398-402 (1995).
17. Wanker, E. et al. HIP-1: a huntingtin interacting protein isolated by the yeast two-hybrid system. *Hum. Mol. Genet.* **3**, 487-495 (1997).
18. Sittler, A. et al. SH3GL3 associates with the Huntingtin exon 1 protein and promotes the formation of polyglu-containing protein aggregates. *Mol. Cell* **4**, 427-436 (1998).
19. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824-7 (2002).
20. Schwikowski, B., Uetz, P. & Fields, S. A network of protein-protein interactions in yeast. *Nature Biotechnol.* **17**, 1030-1032 (2000).
21. Cha, J. H. Transcriptional dysregulation in Huntington's disease. *Trends Neurosci.* **23**, 387-392 (2000).

22. Holbert, S. et al. The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc. Natl Acad. Sci. USA* **98**, 1811-1816 (2001).
23. Faber, P. et al. Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* **7**, 1463-1474 (1998).
24. Hawkes, N. A. et al. Purification and characterization of the human elongator complex. *J. Biol. Chem.* **277**, 3047-3052 (2002).
25. Liu, B., Gross, M., ten Hoeve, J. & Shuai, K. A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. *Proc. Natl Acad. Sci. USA* **98**, 3203-3207 (2001).
26. Sachdev, S. et al. PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev.* **15**, 3088-3103 (2001).
27. Engelender, S. et al. Huntingtin-associated protein 1 (HAP1) interacts with the p150<sup>Glued</sup> subunit of dynactin. *Hum. Mol. Genet.* **6**, 2205- 2212 (1997).
28. Waelter, S. et al. The huntingtin interacting protein HIP1 is a clathrin and alpha-adaptin-binding protein involved in receptor-mediated endocytosis. *Hum. Mol. Genet.* **10**, 1807-1817 (2001).
29. Witke, W. et al. In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J.* **17**, 967-976 (1998).
30. Chook, Y. & Blobel, G. Karyopherins and nuclear import. *Curr. Opin. Struct. Biol.* **11**, 703-715 (2001).
31. Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W. & Bork, P. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1-18 (2001).
32. Wang, T. et al. hADA3 is required for p53 activity. *EMBO J.* **20**, 6404-6413 (2001).
33. Vitale, N. et al. GIT proteins, A novel family of phosphatidylinositol 3,4, 5-trisphosphate-stimulated GTPase-activating proteins for ARF6. *J. Biol. Chem.* **275**, 13901-13916 (2000).
34. Mangiarini, L. et al. Exon 1 of the Huntington's disease gene containing a highly expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493-506 (1996).
35. DiFiglia, M. et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990-1993 (1997).
36. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
37. Bernstein, C., Bernstein, H., Payne, C. M. & Garewal, H. DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat. Res.* **511**, 145-178 (2002).
38. Becher, M. W. et al. Intranuclear neuronal inclusions in Huntington's disease and dentatorubral and pallidoluysian atrophy: correlation between the density of inclusions and *IT15* CAG triplet repeat length. *Neurobiol. Dis.* **4**, 387-397 (1998).
39. Hernandez-Deviez, D. J., Casanova, J. E. & Wilson, J. M. Regulation of dendritic development by the ARF exchange factor ARNO. *Nature Neurosci.* **5**, 623-624 (2002).
40. Di Cesare, A. et al. p95-APP1 links membrane transport to Rac-mediated reorganization of actin. *Nature Cell Biol.* **8**, 521-530 (2000).

41. Zhang, H., Webb, D. J., Asmussen, H. & Horwitz, A. F. Synapse formation is regulated by the signaling adaptor GIT1. *J. Cell. Biol.* **161**, 131-142 (2003).
42. Graveland, G. A., Williams, R. S. & DiFiglia, M. Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. *Science* **227**, 770-773 (1985).
43. Ko, J. et al. Interaction between liprin-alpha and GIT1 is required for AMPA receptor targeting. *J. Neurosci.* **23**, 1667-1677 (2003).
44. Cha, J. H. et al. Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human Huntington disease gene. *Proc. Natl Acad. Sci. USA* **95**, 6480-6485 (1998).
45. Kim, S. et al. The GIT family of proteins forms multimers and associates with the presynaptic cytomatrix protein Piccolo. *J. Biol. Chem.* **278**, 6291-6300 (2003).
46. Scherzinger, E. et al. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl Acad. Sci. USA* **96**, 4604-4609 (1999).
47. Gutekunst, C. A. et al. Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J. Neurosci.* **19**, 2522-2534 (1999).
48. Lunkes, A. et al. Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions. *Mol. Cell* **10**, 259-269 (2002).
49. Davy A., Bello P., Thierry-Mieg N., Vaglio P., Hitti J., Doucette-Stamm L., Thierry-Mieg D., Reboul J., Boulton S., Walhout AJ., Coux O., Vidal M. A protein-protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2001**, Sep; 2(9) : 821-8.

**Claims**

1. A method for generating a network of direct and indirect interaction partners of a disease-related (poly)peptide comprising the steps of
  - (a) contacting a selection of (poly)peptides suspected to contain one or several of said direct or indirect interaction partners with said disease-related (poly)peptides and optionally with known direct or indirect interaction partners of said disease-related (poly)peptide under conditions that allow the interaction between interaction partners to occur;
  - (b) detecting (poly)peptides that interact with said disease-related (poly)peptide or with said known direct or indirect interaction partners of said disease-related (poly)peptide;
  - (c) contacting (poly)peptides detected in step (b) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (b) under conditions that allow the interaction between interaction partners to occur;
  - (d) detecting proteins that interact with said (poly)peptides detected in step (b);
  - (e) contacting said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide, said (poly)peptides detected in steps (b) and (d) and a selection of proteins suspected to contain one or several (poly)peptides interacting with any of the afore mentioned (poly)peptides under conditions that allow the interaction between interaction partners to occur;
  - (f) detecting (poly)peptides that interact with said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide or with said (poly)peptides identified in step (b) or (d); and
  - (g) generating a (poly)peptide-(poly)peptide interaction network of said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide and said (poly)peptides identified in steps (b), (d) and (f).
2. The method of claim 1, wherein said contacting step (e) is effected in an interaction mating two hybrid approach.

3. The method of claim 1 or 2, said method comprising after step (d) and before step (e) the steps of:
  - (d') contacting (poly)peptides detected in step (d) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (d) under conditions that allow the interaction between interaction partners to occur; and
  - (d'') detecting proteins that interact with said (poly)peptides detected in step (d').
4. The method of any one of claims 1 to 3, wherein said disease-related protein is a protein suspected of being a causative agent of a hereditary disease.
5. The method of any one of claims 1 to 4, wherein said disease-related protein is huntingtin and wherein said interaction partners are the interaction partners as shown in tables 1 and 2.
6. The method of any one of claims 1 to 5, said method comprising the step of determining the nucleotide sequence of a nucleic acid molecule encoding a direct or indirect interaction partner of the disease related protein.
7. The method of any one of claims 1 to 6, wherein said selections of proteins are translated from a nucleic acid library.
8. The method of any one of claims 1 to 7, wherein said selection of proteins in step (a) and/or (c) and/or (d') and/or (e) is the same selection or a selection from the same source.
9. The method of any one of claims 1 to 7, wherein said selection of proteins in step (a) and/or (c) and/or (d') and/or (e) is a different selection or a selection from a different source.
10. The method of any one of claims 1 to 9, wherein said method is performed by contacting the proteins on an array.

11. The method of any one of claims 1 to 10, wherein said interactions are detected by using the yeast two-hybrid system.
12. The method of any one of claims 1 to 11, containing after step (b), (d), (d'') or (f) the additional steps of isolating a nucleic acid molecule with homology to said cDNA expressing the encoded protein and testing it for its activity as a modulator of huntingtin, wherein said nucleic acid molecule is DNA, or RNA, preferably cDNA, or genomic or synthetic DNA or mRNA.
13. A nucleic acid molecule encoding a modulator of huntingtin, wherein said modulator is a protein selected from table 3.
14. The nucleic acid molecule of claim 13, wherein said nucleic acid molecule is DNA, preferably cDNA, genomic DNA, or synthetic DNA or RNA, preferably mRNA.
15. The nucleic acid molecule of claim 13 or 14 fused to a heterologous nucleic acid molecule.
16. The nucleic acid molecule of claim 15, wherein the heterologous nucleic acid molecule encodes a heterologous (poly)peptide.
17. A vector comprising the nucleic acid molecule of any one of claims 13 to 16.
18. A host cell containing the nucleic acid molecule of any one of claims 13 to 16 or the vector of claim 17.
19. A method of producing a (poly)peptide, comprising culturing the host cell of claim 18 under conditions such that the (poly)peptide encoded by said polynucleotide is expressed and recovering said (poly)peptide.
20. A (poly)peptide comprising an amino acid sequence encoded by a nucleic acid molecule of any one of claims 13 to 16, or which is chemically synthesized, or is obtainable from the host cell of claim 18, or which is obtainable by the method of claim 18.

21. The (poly)peptide of claim 20 fused to a heterologous (poly)peptide.
22. A protein complex comprising at least two proteins, wherein the two proteins are selected from the group of interaction partners listed in table 4.
23. An antibody specifically recognizing the (poly)peptide of claim 20 or 21 or specifically reacting with the protein complex of claim 22.
24. The antibody of claim 23 which is polyclonal, monoclonal, chimeric, single chain, single chain Fv, human antibody, humanized antibody, or Fab fragment.
25. A method of identifying whether a protein promotes huntingtin aggregation, comprising
  - (a) transfecting a first cell with a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates;
  - (b) co-transfected a second cell with
    - (i.) a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates; and
    - (ii.) a nucleic acid molecule encoding a candidate modulator protein identified by the methods of any one of claims 1 to 12 or a nucleic acid molecule encoding a modulator protein selected from table 1 or table 2;
  - (c) expressing the proteins encoded by the transfected nucleic acid molecule of (a) and (b);
  - (d) isolating insoluble aggregates of huntingtin from the transfected cell of (a) and (b); and
  - (e) determining the amount of insoluble huntingtin aggregates from the transfected cell of (a) and (b)wherein an increased amount of huntingtin aggregates isolated from the transfected cells of (b) in comparison with the amount of huntingtin aggregates isolated from the transfected cells of (a) is indicative of a protein's activity as an enhancer of huntingtin aggregation.

26. A method of identifying whether a protein inhibits huntingtin aggregation, comprising
- (a) transfecting a first cell with a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates;
  - (b) co-transfected a second cell with
    - (i.) a nucleic acid molecule encoding a variant of the huntingtin protein or a fragment thereof capable of forming huntingtin aggregates ; and
    - (ii.) a nucleic acid molecule encoding a candidate modulator protein identified by the methods of any one of claims 1 to 12 or a nucleic acid molecule encoding a modulator protein selected from table 1 or table 2;
  - (c) expressing the proteins encoded by the transfected nucleic acid molecule of (a) and (b);
  - (d) isolating insoluble aggregates of huntingtin from the transfected cell of (a) and (b); and
  - (e) determining the amount of insoluble huntingtin aggregates from the transfected cell of (a) and (b)  
wherein a reduced amount of huntingtin aggregates isolated from the transfected cells of (b) in comparison with the amount of huntingtin aggregates isolated from the transfected cells of (a) is indicative of a protein's activity as an inhibitor of huntingtin aggregation.
27. The method of claim 25 or 26, wherein prior to step (d) the cells are treated with an ionic detergent.
28. The method of any one of claims 25 to 27, wherein the huntingtin aggregates are filtered onto a membrane.
29. A method for identifying compounds affecting an interaction of huntingtin or of a direct or indirect interaction partner of huntingtin comprising

- (a) contacting interacting proteins selected from the group of interacting proteins listed in table 1 and/or table 2 in the presence or absence of an potential modular of interaction;
  - (b) identifying compounds capable of modulating said interaction.
30. The method of any one of claims 25 to 29 , further comprising
- (a) modeling said compound by peptidomantics and
  - (b) chemically synthesizing the modeled compound.
31. The method of any one of claims 25 to 30, wherein said compound is further modified to achieve
- (i) modified site of action, spectrum of activity, organ specificity, and/or
  - (ii) improved potency, and/or
  - (iii) decreased toxicity (improved therapeutic index), and/or
  - (iv) decreased side effects, and/or
  - (v) modified onset of therapeutic action, duration of effect, and/or
  - (vi) modified pharmokinetic parameters (resorption, distribution, metabolism and excretion), and/or
  - (vii) modified physico-chemical parameters (solubility, hygroscopicity, color, taste, odor, stability, state), and/or
  - (viii) improved general specificity, organ/tissue specificity, and/or
  - (ix) optimized application form and route
- by
- (i) esterification of carboxyl groups, or
  - (ii) esterification of hydroxyl groups with carbon acids, or
  - (iii) esterification of hydroxyl groups to, e.g. phosphates, pyrophosphates or sulfates or hemi succinates, or
  - (iv) formation of pharmaceutically acceptable salts, or
  - (v) formation of pharmaceutically acceptable complexes, or
  - (vi) synthesis of pharmacologically active polymers, or
  - (vii) introduction of hydrophilic moieties, or
  - (viii) introduction/exchange of substituents on aromates or side chains, change of substituent pattern, or
  - (ix) modification by introduction of isosteric or bioisosteric moieties, or

- (x) synthesis of homologous compounds, or
  - (xi) introduction of branched side chains, or
  - (xii) conversion of alkyl substituents to cyclic analogues, or
  - (xiii) derivatisation of hydroxyl group to ketales, acetals, or
  - (xiv) N-acetylation to amides, phenylcarbamates, or
  - (xv) synthesis of Mannich bases, imines, or
  - (xvi) transformation of ketones or aldehydes to Schiff's bases, oximes, acetals, ketales, enoesters, oxazolidines, thiazolidines or combinations thereof.
32. A method of diagnosing Huntington's disease in a biological sample comprising the steps of
- (a) contacting the sample with an antibody specific for a protein of table 1 or 2 or an antibody specific for the protein complex of claim 22; and
  - (b) detecting binding of the antibody to a protein complex, wherein the detection of binding is indicative of Huntington's disease or of a predisposition to develop Huntington's disease.
33. The method of claim 32, wherein
- (a) said protein complex contains GIT1 or
  - (b) said antibody is specific for a protein complex containing GIT1.
34. A diagnostic agent/composition or pharmaceutical composition comprising the nucleic acid molecule of any one of claims 13 to 16, the (poly)peptide of claim 20 or 21 or the (poly)peptide mentioned in anyone of tables 1 and 2, the antibody of claim 23 or 24, an antibody specifically reacting with a protein selected from table 2 and/or a protein selected from table 2.
35. Use of the molecule of any one of claims 13 to 16, the (poly)peptide of claim 20 or 21 or the (poly)peptide mentioned in anyone of tables 1 and 2, the antibody of claim 23 or 24, an antibody specifically reacting with a protein selected from table 2 and/or a protein selected from table 2, for the preparation of a pharmaceutical composition for the treatment of Huntington's disease.

20. Juni 2003

**- ABSTRACT**

The present invention relates to a method for generating a network of direct and indirect interaction partners of a disease-related (poly)peptide comprising the steps of (a) contacting a selection of (poly)peptides suspected to contain one or several of said direct or indirect interaction partners with said disease-related (poly)peptides and optionally with known direct or indirect interaction partners of said disease-related (poly)peptide under conditions that allow the interaction between interaction partners to occur; (b) detecting (poly)peptides that interact with said disease-related (poly)peptide or with said known direct or indirect interaction partners of said disease-related (poly)peptide; (c) contacting (poly)peptides detected in step (b) with a selection of (poly)peptides suspected to contain one or several (poly)peptides interacting with said (poly)peptides detected in step (b) under conditions that allow the interaction between interaction partners to occur; (d) detecting proteins that interact with said (poly)peptides detected in step (b); (e) contacting said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide, said (poly)peptides detected in steps (b) and (d) and a selection of proteins suspected to contain one or several (poly)peptides interacting with any of the afore mentioned (poly)peptides under conditions that allow the interaction between interaction partners to occur; (f) detecting (poly)peptides that interact with said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide or with said (poly)peptides identified in step (b) or (d); and (g) generating a (poly)peptide-(poly)peptide interaction network of said disease-related (poly)peptide and optionally said known direct or indirect interaction partners of said disease-related (poly)peptide and said (poly)peptides identified in steps (b), (d) and (f). Moreover, the present invention relates to a protein complex comprising at least two proteins and to methods for identifying compounds interfering with an interaction of said proteins. Finally, the present invention relates to a pharmaceutical composition and to the use of compounds identified by the present invention for the preparation of a pharmaceutical composition for the treatment of Huntington's disease.

EPO - Munich  
3  
20. Juni 2003

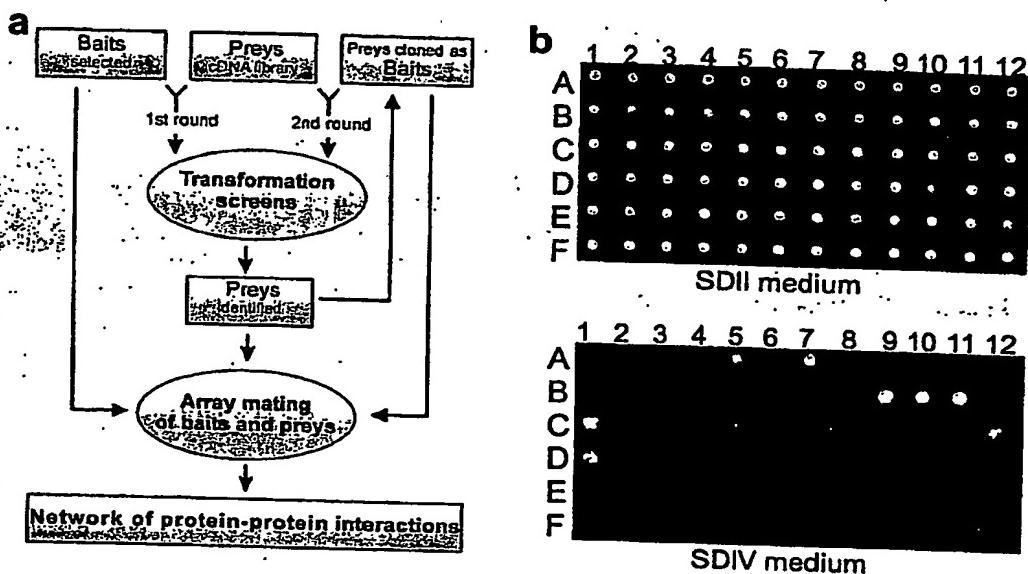


Figure 1

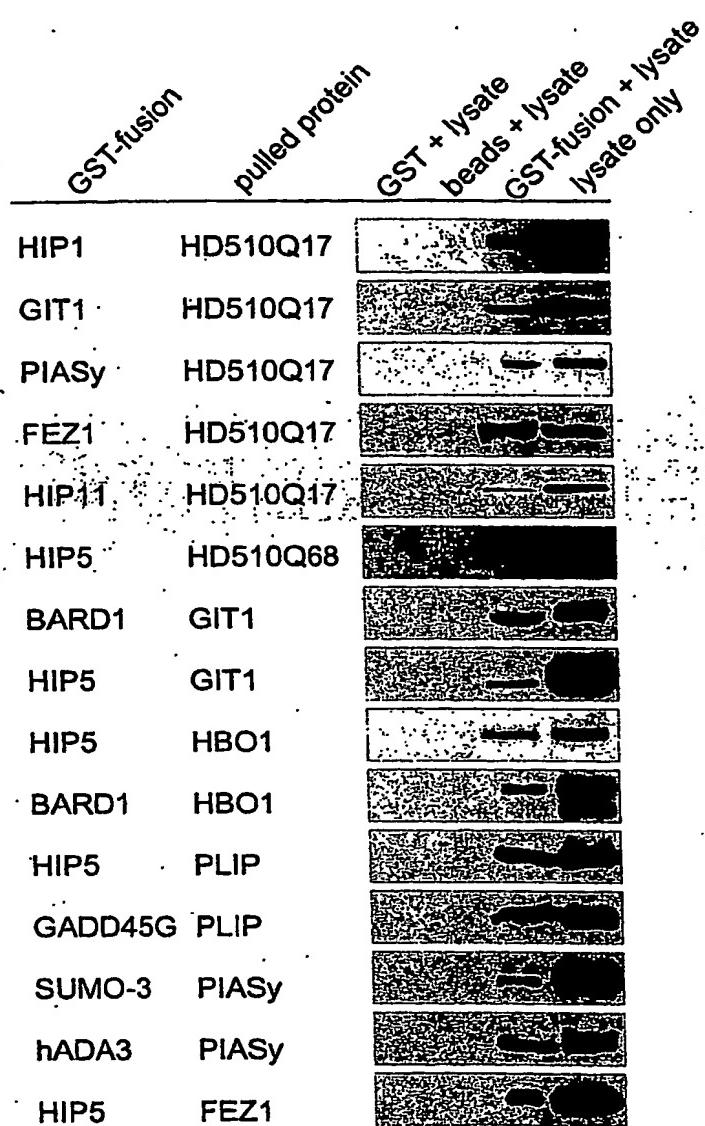


Figure 3

>ALEX2  
AESVVGAAAMASAIAPPGVTEALGAEAPAMAGAPKVAEAPREAEATSRAAVPPGTVVPTTEAAAPTE  
VTEGPVAAPTKVAEAPGVASPTEAAEAPVPATPTGAAAPTGAESPGTSGSPRTAVVPGTSAAKK  
ATPGAHTGAIPKATSATGAVPKGGKGVTRSRNGGKGKGSKSKVEVDELMGFRRPGDGAACAAAAS  
ANGGQAFLAEVPDSEEGESGWTDTESDSDSEPETQRRGRGRRPVAMQKRPPYEIDEILGVRDLRK  
VLALLQKSDDPFIQQVALLTLSNNANSCNQETIRKLGLLPIIANMINKTDPHIKEKALMAMNNLS  
ENYENQGRLQVYMNKVMDIMASNLNSAVQVUGLKFLTNMTITNDYQHLLVNSIANFRLLSQGGG  
KIKVEILKILSNFAENPDMLKKLLSTQVPASFSSLYNSYVESBILINALTLFEIYDNLRAEVFN  
REFNKGSFLYLCCTSGVCVKIRALANHDLVVKVVIKLVNKF  
>APP1  
EEEEESFPQPVDDYFVEPPQAEAAAETVPPPSHTLAUVGKVTPTPRTDGVDIYFGMPGEISHE  
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLKPADROALNEHFQSILQTLLEEQVSGERQROLVE  
THATRIVALINDQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAADVPE  
KAQQMRFQVHILQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPPGGSSED  
KGGLQPDSDSKDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKVNASVPGVSLSTHRRFRGMSWHQ  
LGQGCPVRLCRVC  
>BAIP1  
RPRTKMATAMYLEHYLDSIENLPCELQRNFQLMRELDQRTEDKKAEIDLAAEYISTVKTLSPDQR  
VERLQKIQNAYSCKKEYSDDKVQLAMQTYEMVDKHIRRLDADLARFEADLKDKMEGSDFESSGGRG  
LKKGRGQKEKRGSRGRRTSEEDTPKKKKHKGSEFTDTILSVHPSDVLDMPVDPNEPTYCLCHQ  
VSYGEMIGCDNPDCPIEWFHACVDLTTKPKGKW  
>BAIP2  
SQQASVTMHDVDAESFEVLVDYCYTGRVSLSEANQRLYAASDMLQLEYVREACASFLARRLDLT  
CTAILKFADAFDHKLRSQAQSYIAHNFKQLSRMGSIREETLADLTLAQLLAVRLRDSLIESERT  
VCHVAVQWLEAAAKERGPSAAEVFKCVRWMHFTeedQDYLEGLLTkpIVKKYCLDVIEGALQMRYG  
DILYKSLVPVPNSSSSSSSNSLVSAAEPPQRLGMCAKEMVIFFGHPRDPFLCYDPYSGDIYTMP  
SPLTSFAHTKTVT&SAVCVSPDHDIYLAAQPRKDlwVYKPAQNSWQQLADRLLCREGMDVAYLNGY  
IYILGGRDPITGVKLKEVECYSVQRNQWALVAPVPHSFYSELIVQNYLYAVNSKRMLCYDPSHN  
MWLNCAASLKRSDFQEACVFNDEIYCICIDIPVMKVYNPARGEWRRISNIPLDSETHNYQIVNHQKL  
LLITSTTPQWKKNRVTVYEYDTREDQWINIGTMGLLQFDSGFICLCARVYPSCLEPGQSFITEED  
DARSESSTEWDLDGFSELDSESQSSFSDDDEVVQAPQRNAQDQQQGSL  
>BAIP3  
GHNAPRKVTAVIYARKGSVILQSIKISSVDATTVTSSQCVFRDQEPRKIHNEMASTSDKGAQGRND  
KKDSQGRSNKALHLKSDAEFFKKIFGLTKDLRVCLTRIPDHLTSGEGFDSFSSLVKSGTYKETEFMV  
KEGERKQQNFDKKRKAKTNKMDHIKKRKTENAYNAIINGEANVTGSQOLLSSILPTSDVSQHNILT  
SHSKTRQEKRTEMYYTHEKQEKGTLNSNAAYEQSHFFNKNYTEDIFPVTPPELEETIRDEKIRRL  
KQVLREKEAALEEMRKKMHQK  
>BARD1  
LAGFESLTCSPVVSRGLLASRSPRSLSSEGGIMPNDNRQPRNRQPRIRSGNEPRSASAMEPDGRGA  
WAHSRAALDRLEKLLRCRCTNILREPVCGLGCEHIFCSNCVSDCIGTCPVYTPAWIQLDKINR  
QLDSMIQLCSKLRNLLHDNEPSDLKEDKPRKSLFNDAGNKKNSIKMWFSRSGKVKRYVVKASVQT  
QPAIKKDASAQQDSYEFVSPSPPADVSEAKASARSGKKQKKKTLABINQKWNLEAEKEDGEFDS  
KEESQKLVFSFCQPSVISSPQINGEIDLILLASGSLTESECFCGSLTEVSLPLAEQIESPDTKSNEV  
UTPEVKCKNYLTSKKSLPLENNNGKRGHHNRLLSPISKRCRTSILSTSGDFVKQTVPSENIPLPECS  
BPPSCKRKVGGTSGSKTVTCPMNSLVFHQVHLLH  
>CA150  
QQFIPGPLKILVWPCCFLSQAPTTQDQTPSSAVSVATPTSVSTPAPTAATPVQTVQPHPQTLPPA  
VPHSVPQPTTAIPAFPPVMPPFRVLPGMPILPLPGVLPGMAPPIVPMIHQVIAIAASPATLAGAT  
AVSEWTEYKTADGKTYYYNNRTLESTWEKPQELKEKEKLEEKIKEPIKEPSEEPLPMETEEEDPKE  
EPIKEIKEPKEEEMTEEKAQAKPVATAPIPGTPWCWVWTGDERVFYNTTRLSMWDRPDDL  
IGRADVDKIIQEPPHKKGMEELKKLRHPTPTMLSIQKWFQSMSAIKEEQELMSEINEDEPVKAKKR

**Figure 6**

&gt;CGI-125

FDASARNFARVSGLLL CQAGGVLVSSFVMAAAVAMETDDAGNRLRFQLELEFVQCLANPNYLNFLA  
 QRGYFKDKAFVNLYKLLYWKDPEYAKYLKPQCLHMLLELLOYEHFRKELVNAQCACKFIDEOQQILH  
 WQHYSRKRMRLQQALAEQQQQNNTSGK

&gt;CGI-74

VEKARAKKREAAEVYRNSMPASSFQQQKLRVCEVCSAYLGLHDNDRRLADHFGGKLHLGFIEIREK  
 LEEILKRVVAEKQEKRNQERLKRREREREREKLRRSRSHSKNPKR

&gt;CLH-17

MAQILPIRFQEHLQLQNLGINPANIGFSTLTMESDKFICIREKVGEQAQVVIIDMNDPSNPIRRPI  
 SADS AIMNPASKVIALKAGKTLQIFNIEMSKMKHAHTMTDDVTFWKWISLNTVALVTDNAVYHWSM  
 EGESQPVVKMFDRHSSLAGCQIINYRTDAKQKWLTTGISAQQNRVVGAMQLYSVDRKVSQPIEGHA  
 ASFAQFKMEGNAEESTLFCFAVRGQAGGKLHIIIEVGTPTGNQFPKKAVDVFFPPEAQNDFPVAM  
 QISEKHDVVFLITKYGYIHLYDLET

&gt;CLK1

DAWVLEHLNTTDPNSTFRCVQMLEWFEHHGHICIVFELLGLSTYDFIKENGFLPFRLDHIRK MAYQ  
 ICKSVNFLHSNKLTHDLKPENILFVQSDYTEAYNPKIKRDERTLINPDIKVVDFGSATYDDEHHS  
 TLVSTRHYRAPEVILALGWSPQPCDVWSIGCILIEYYLGFVFTPHDSKEHLAGMERILGPLPKHMI  
 QKTRKRKYFHDLWDHRSAGRYVSRRCKPLKEFMLSQDVHERLFDLIQKMLEYDPAKRITLR  
 EALKHPFFDLLKKSI

&gt;DRP-1

KDNFTLIPGVNGIEERMTVVWDKAVATGKMDENQFVAVTSTNAAKIFNLYPRKGRIAVGSDADVV  
 IWDPDKLKTITAKSHKSAVEYNIFEGMECHGSPLVVISQGKIVFEDGNINVNKGMGRFIPRKAFPE  
 HLYQRVKIRNKVFGLQGVSRGMYDGPVYEVPATPKYATPAPSAKSSPSKHOPPPIRNLHQSNFSLS  
 GAQIDDNNPRRTGHRIVAPPGRSNITSLG

&gt;EF1A

MHHEALSEALPGDNVGFNVKNVSVKDVRGNVAGDSKNDPPMEAAGFTAQVIIILNHPGQISAGYAP  
 VLDCHTAHIACKFAELKEKIDRRSGKKLEDGPKFALKSGDAIAIVDMVPGKPMCVESFSDYPPLGRFA  
 VRDMRQTAVGVVIKAVDKKAAGAGKVTSAQKAQKAK

&gt;EF1G (bait)

AAGTLYTYPEWRAFKALIAAQSGAQVRVL SAPPHFHGQTNRTPFELRKFPACKVPAFEGDDGF  
 CVFESNAIAYYVSNEELRGSTPEAAAQVWVSPFADSDIVPPASTWVFTLGIMHHNKQATENAKE  
 EVRILGLLDAYLKTRTFLVGERVTLADI TVVCTLWLYKQVLEPSFRQAFPNTNRWF LTCINQPQ  
 FRAVLGEVKLCEKMAQFDAKKFAETQPKKDTPRKEGSREEKQKQPAERKEEKAAAAPAPEEEMDE  
 CEQALAAEPKAKDPFAHLPKSTFVLDDEFKRKYSNETTLSVALPYFWEHFDKDGWSLWYSEYRFPEE  
 LTQTFMSCNLITGMFQRLDKLRKNAFASVILFGTNSSSISGVWVFRGQELAFPLSPDWQVDYESY  
 TWRKLDPGSEETQTLVREYFSWEGAFQHVGKAFNQGKIFK

&gt;EF1G (prey)

AAGTLYTYPEWRAFKALIAAQSGAQVRVL SAPPHFHGQTNRTPFELRKFPACKVPAFEGDDGF  
 CVFESNAIAYYVSNEELRGSTPEAAAQVWVSPFADSDIVPPASTWVFTLGIMHHNKQATENAKE  
 EVRILGLLDAYLKTRTFLVGERVTLADI TVVCTLWLYKQVLEPSFRQAFPNTNRWF LTCINQPQ  
 FRAVLGEVKLCEKMAQFDAKKFAETQPKKDTPRKEGSREEKQKQPAERKEEKAAAAPAPEEEMDE  
 CEQALAAEPKAKDPFAHLPKSTFVLDDEFKRKYSNETTLSVALPYFWEHFDKDGWSLWYSEYRFPEE  
 LTQTFMSCNLITGMFQRLDKLRKNAFASVILFGTNSSSISGVWVFRGQELAFPLSPDWQVDYESY  
 TWRKLDPGSEETQTLVREYFSWEGAFQHVGKAFNQGKIFK

&gt;FEZ1

GNCSDTEIHEKEEEBFNEKSENDSGINEEPLLTAQVIEEIEEMMQNSPDPEEEEVLEEDGET  
 SSQADSVLLOEMQALTQTFNNNWSYEGLRHMSGSELTELDDQVEGAIRDSEELVQQLARRDELF  
 EKEVKNSFITVLIIEVQNQKQEORELMKKRRKEGLSLQSSRIEKGNQMLKRFMSMEGI.SNILQSGI  
 RQTFGSSGTDKQYLNTVIPYEKKASPPSVEDLQMLTNILFAMKEDNEKVPTLLTDYILKVL.CPT

Figure 6 (continued)

>G45IP1  
MASSGGELGSLFDHHVQRAVCDTRAKYREGRRPRAVKVYTINLESQYLLIQGVPAVGVMKELVERF  
ALYGAIEQYNALDEYPAEDFTEVYLIKPMNLQSARTAKRMDEQSFVGGLLHVCYAPEFETVEETR  
KKLQMRKAYVVKTENKOHYVTKKLVTEHKDTEFRQDFHSEMSGFCKAALNTSAGNSNPYLPYS  
CELPLCYFSSKCMCSSGGPVDRAPDSSKGDRNHHKTMGHYNHNDSLRKTOINSLKNSVACPGAQKA  
ITSSEAVDRFMPRTTQLQERKRRREDDRKLGTFLQTNPTGNEIMIGPLLPDISKVDMHDDSLNTTA  
NLIRHKLKEVISSVPKPPEDKPEDVHTSHPLKQRRRI  
>G45IP2  
RTCMPYIFSLSLEALKCPRIRNNEKMLSDSHGVETIRDILPDTSLGGSFFKIIITAKAVLKLQAGN  
AEEAALWRDLVRKVLASYLETAAEAVTLGGSDLDENCOEVLFATRENGFLQYLVAIPMEKGLDQ  
GCFCAGCSRQIGFSFVRPKLCAPSGLYCDICHQDDASVI PARI IHNWDLTKRPICRQALKFLTQI  
RAQPLINLQMVNASLYEHVERMHLIGRRREQLKLGDYLGLCRSGALKEALKRLNHRNYLLESPHR  
FSVADLQQIADGVYEGFLKALIEFASQHVVYHCDLCTQRGFICQICQHHDIIIFPFEFDTTVRCAECK  
TVFHQSCQAVVKKGCPRCARRKYQEQNIFA  
>G45IP3  
PNRGPLSPNDLRPSHVISLPLHNAPHTRPTNQHTNHIPMMARCNTRKHIPRPPHTTCPKRPSIRD  
NPIYYLRSFFLRRIFLSLLPLQSPYPPIRRALAPNRHHPAKSPRSPTPKHIRITRIRSINHLSSP  
>GADD45G  
GAGAEPGLECGWSWGAKVCRWPGLSPRPPAGSRSLRWLLRRMQGAGKALHELLL SAQRQGCLT  
AGVYESAKVLNVDPDNVTFCVLAAGEEDEGDI ALQIHFTL IQAFCCENDIDIVRVGDVQLAAIVG  
AGEEAGAPGDLHCILI SNPNEAWKDPALEKLSLFCESRSVNDWVPSITLPE  
>GIT1  
PQADRSRQKCMSQSLDSELAKAKKLQALSNRLFEELAMDVYDEVDRRENDAVWLATQNHSTL  
VTERSAVPFLPVNPEYSATRNQGRQKLARFNAREFATLI IDILSEAKRQQGKSLSSPTDNLELSL  
RSQSDLDDQHDYDSVASDEDTDQEPLRSTGATRSNRARSMDSSLSDGAVTLOEYLELKALATSE  
AKVQQLMKVNSSLSDELRLQREIHKLQAENLQLRQPPGPVTPPLPSERAETHPMAPGGSTHRD  
RQAFSMYEPGSALKPFGPPGDELTTRLQPFHSTELEDDAIYSVHVPAGLYRIRKGVSASAVPFTP  
SSPLLSCSQEGSRHTSKLSRHGSGADS DVENTQSGDPLLGLEGKRFLELGKEEDFHP ELES LDGDL  
DPGLPSTEDVILKTEQVTKNIQELLRAAQEFKHDSFVPCSEKIHLAVTEMASLFPKRPALEPV RSS  
LRLINASAYRLQSECRTVPPPEPGAPVDFQLLTQVIQCAYDIAKAAKQLVTIT TREKKQ  
>hADA3  
KDVDALLKKSEAQHEQPEDGCPFGALTQRLLQALVEENI ISPMEDSP IPDMSGKESGADGASTSPR  
NQNKPFSVPHTKSLESRIKEELIAQGLLESED RPAEDSEDEVLAELRKROAEELKALSAHNRTKKHD  
LLRLAKEEVSRQELRQVRVMADNEVMDA FRKIMAARQKKRPTKKEKDQAWKTLKERESILKLLDG  
>HB01  
DAERQEALGIVRRIGTDTEAATEPAGATVPAAAAARIGTVGPQPPAMPRKRNAGSSSDGTEDSD  
FSTDLEHTDSSESDGT SRRSARVTRSSARL SQSSQDSSPVRNLQSF GTEEPAYSTRRVTRSQQOPT  
PVTPKKYPLRQTRSSGSETEQVVD FSDRET KNTADHDES PRTPTGNAPSSES DIDISSPNVSHDE  
SIAKDMSLKDSGSDLSHPKRRFHESYNFNMKCPTPGCNSLGHLTGKHERHFSISGCPLYHNLSA  
DECKVRAQS RDQIEERMLSHRQDDNNRHATRHQAPTER QLRYKEKVAELRKKRNSGLSKEQKEKY  
MEHQTYGNTREPLLENLTSEYDLDLFRAQARASEDLEKLR LQGQITEGSNMIKTI AFGRYELDT  
WYHSPYPEEYARLGRILYMC EFCLKYMKSQTILRRHMAKCVWKHPPGDEIYRKGSISVFEVDGKKNK  
IYCQNLCLLAKLFLDHKTLYYDVEPFLFYVMTEADNTGCHLIGYFSKEKNSFLNVNVSCILT MPQY  
MRQGYGKMLIDFSYLLSKVEEKVGS PERLSDLGLISYRSWKEVLLRYLHN FQGKEISIKEISQE  
TAVNPVDIVSTLQALQMLKYWKKGKHLVLKRQDLIDEWI AKEAKRSNSNKTMDPSCLKWTPPKGT

Figure 6 (continued)

>HD1.7

MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPQLPQPPPQAQPLLPQ  
PQPPPPPPPPPGPAVAEEPLHRPKELSATKKDRVNHCCLTICENIVAQSVRNSPEFKQLLGIAME  
LFLLCSDDAESDVRMVADECILNKVIKALMDSNLPRLQLELYKEIKKNGAPRSLLRAALWRFAELAHL  
VRPQKCRPYLVNLLPCLTRTSKRPEESVQETLAAAVPKIMASFGMFANDNEIKVLLKAFIANLKSS  
SPTIRRTAAGSAVSICQHSRRTQYFYSWLLNVLLGLLVPVEDEHSTLLILGVLLTRYLVPPLLQQQ  
VKDTSLKGSGVTRKEMEVSPSAEQLVQVYELTLHHTQHQDHNVVTGALELLQQLFRTPPPPELLQT  
LTAVGGIGQLTAKEESGGRSRSGSIVELIAGGGSSCPVLSRKQKGKVLLGEEEALEDDESRSRD  
VSSSALTASVKDEISGELAASSGVSTPGSAGHDIITEQPRSQHTAGGLSGSGQL

>HDD1.0

MATLEKLMKAESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPPPOLPQPPPQAQPLLPQ  
PQPPPPPPPPPGPAVAEPLHRPKELSATKKDRVNHCLTICENIVAQSVRNSPEFKQLLGIA  
LFLLCSDDAESDVRMVADECLNKVIKALMDNSLPRLQLELYKEIKKNGAPRSLRAALWRFAELAHL  
VRPQKCRPYLVNLLPCLTRTSKRPEEVQETLAAAVPKIMASFGNFANDNEIKVLLKAFIANLKSS  
SPTIRRTAAGSAVSICQHSRTQYFYSWLLNVLLGLLVPVEDEHSTLLILGVLLTLSR

>HDd1.3

PRLQLELYKEIKKNGAPRSLRAALWRFAELAHLVRPQKCRPYLVNLLPCLTRTSKRPEESVQETLA  
AAVPKIMASFGNFANDNEIKVLLKAFIANLKSSSPTIRRTAAGSAVSICQHSRTQFYFWLNLNVL  
LGLLVPUVEDEHSTLLILGVLLTLRYLVPPLLQQVKDTSLKGSFGVTRKEMEVSPSAEQLVQVYELT  
LHHTQHQDHNVVTGALELLQLFRTPPPPELLQTLTAVGGIGQLTAAKEESGGRSRSGSIVELIAGG  
GSSCSPVLSRKQKGKVLLGEEEALEDDSESRSVDSSALTASVKDEISGELAASSGVSTPGSAGHD  
IIITEQPRSQHTAGGLSGSGQL

>HDexQ20

MATLEKMMKA FESLKS FQQQQQQQQQQQQQQQQQQQ P PPPPPP QLPQPQQAQPLLQP  
OPPPPPPPPPPGPAVAEEPLHRP

>HDex051

>HIP1

ADTLQGHRDRFMEQFTKLKDLFYRSSNLQYFKRLIQIPQLPENPPNFLRASALSEHISPVVVPAE  
ASSPDSEPVLEKDDLMDMDASQQNLFDNKFDDIFGSSFSDPFNFSQNGVNKDEKDHLLIERLYRE  
ISGLKAQLENMKTESQRVVLQLKGHVSELEADLAEQQHLRQQAADDCEFLRAELDELRRQREDTEK  
AQRSLSEIERKAQANEQRYSKLEKYSELVONHADLLRKNAEVTKQVSMARQAQVDLEREKKELED  
SLERISDQGQRKTQEOLLEVLESLKQELATSQRELQVILQGSLETSAQSEANWAAEFIALEKERDSLVS  
SGAAHREEELSALRKELDTQKLASTEESMCOLAKDORKMLLVGSRKAAEVOIODASTRB

>HIP11

VDLVTACDIRYCAQDAFFQVKEVDVGLAADVGTQLQRLPKVIGNQSLVNELAFTARKMMADEALGSG  
LVSRVFPDKEVMILDAALALAAEISSKSPVAVQSTKVNLLYSRDHSAESLNVASWNMSMLQTQDL  
VKSVOATTENKEI.KTVTESKT.

>HTP13

PCCSED TIPSQVSDYDYFSVSGDQEADQQEFDKSSTIPRNSDIQS SYRRMFQAKRPASTAGLPTTL  
GPAMVTPGVATIRRTPSTKPSVRRGTIGAGPIPIKTPVIPVKTPVDPDLPGVLPAPPDGPEERGEH  
SPESPSPVGEGPQGVTSMPSSMWSGQASVNPPLPGPKPSPPEEHROQIPESEAEQEREPPSATVSP  
GQIPESDPADLSPRDT PQGEDMLNAIRGVKLKKTTNDRSAPRFS

>HIP15

IHMAPPNLNMIEFTICQVCEETLAHSVDSLEQLTGIRMLRHTMTIDYHTLIANYMSGFLSLLT  
TANARTKFHVLKMLLNLSENPAVAKKLFSAKALSIFVGFLFNIETNDNIQIVIKMFQNISNIKSG  
KMSLIDDDFSLEPLISAFREFEELAKOLOAQIDNONDPEVGOOS

**Figure 6 (continued)**

>HIP16  
DEEERNHRQMIKEAFAGDDVIRDFLKEKREAVEASKPKDVLTLPGWGEWGVGLKPSAKKRRFL  
IKAPEGPPRKDKNLPNVIINEKRNIHAAAHQVRVLPPFTHWQFERTIQTPIGSTWNTQRAFQKL  
TTPKVVTKPGHIINPIKAEDVGYRSSRSDLSVIQRNPKRITTRHKKQLKKCSVD

>HIP2  
MANIAVQRIKREFKEVLKSEETSKNQIKVDLVDEDNFTELRGEIAGPPDTPYEGGRYQLEIKIPETY  
PFNPPKVRFITKIWHPNISSVTGAIICLDILKDQWAAMTLRTVLLSLQALLAAEPDDPQDAVVAN  
QYKQNPEMFQOTARLWAHVVAGAPVSSPEYTKKIELCAMGFDRNAVIVALSSKSWDVETATELL  
SNX

>HIP5 (bait)  
FLKSILKESKYEHGYLKALIINOSFKFGNQKAAIRDSIELTKEKGAEIPKTIKKLRWFDETSNI  
ENNAENSHSLKNKTGTQQHSQQFHIQSGAGSNIISVSTCAVNSADTKKSREDSISENVTTLGGSG  
ADHMPLNCIFPSGYNFAKHAWPASKKEESKIPVHDDSKTKQGPQRGRAKIIRKPGSAKVQSGFIC  
TNRKGAVIOPOSASKVNIFTQAQGKLIIPCPPPQSTSNIIRSGKNIQVSQCQPVTPENPQNIITHNS  
FNSKHVLPTEHSLNQWNQESSPLSNACSDLTVIPSLPSYCSECQTFAKINHSNGTQAVARQDA  
TLYCTQRSPVCEESYPSVTLRTAEEEESVPLWKRGPNVLHQNKRTGSTVMRRKRIAETKRRNILEQ  
KRQNPQGSVGQKYSEQINNFGQSVLLSSSEPQKTRGTSYIEEVSDSTSEFLMAENLVKAQPEDEI  
LTVLNSKQIQKSNLPLNKTOQFNICTLSAEEQKILESLNDLNERLHYIQESICKNPSIKNTLQIIP  
LLEKREDRTSSCRDKR

>HIP5 (prey)  
FLKSILKESKYEHGYLKALIINOSFKFGNQKAAIRDSIELTKEKGAEIPKTIKKLRWFDETSNI  
ENNAENSHSLKNKTGTQQHSQQFHIQSGAGSNIISVSTCAVNSADTKKSREDSISENVTTLGGSG  
ADHMPLNCIFPSGYNFAKHAWPASKKEESKIPVHDDSKTKQGPQRGRAKIIRKPGSAKVQSGFIC  
TNRKGAVIOPQSASKVNIFTQAQGKLIIPCPPPQSTSNIIRSGKNIQVSQCQPVTPENPQNIITHNS  
FNSKHVLPTEHSLNQWNQESSPLSNACSDLTVIPSLPSYCSECQTFAKINHSNGTQAVARQDA  
TLYCTQRSPVCEESYPSVTLRTAEEEESVPLWKRGPNVLHQNKRTGSTVMRRKRIAETKRRNILEQ  
KRQNPQGSVGQKYSEQINNFGQSVLLSSSEPQKTRGTSYIEEVSDSTSEFLMAENLVKAQPEDEI  
LTVLNSKQIQKSNLPLNKTOQFNICTLSAEEQKILESLNDLNERLHYIQESICKNPSIKNTLQIIP  
LLEKREDRTSSCRDKR

>HMP  
QEQUKIESLAKSLEDALRQTAUTLQAIAAQNAAVQAVNAHSNILKAAMDNSEIAGEKKSQWRTV  
EGALKERRKAVDEAADALLKAKEELEMKSVIENAKKEVAGAKPHITAAEGKLHNIVDLDNVVK  
KVQAAQSEAKVVSQYHELVVQARDDFKRELDSTPEVLPGWKGMSVSDLADKLSTDDLNSLIAAH  
RRIDQLNRELAEQKATEBKQHITLALEKOKLEEKRAFDAVAKALEHIRSEIQAEQDRKIEEVRDAM  
ENEMRTQLRQAAAHTDHLRDVLRVQEQLSEKFEONLSEKLFQRLSQEQVDNFITLDINT  
AYARLRGIEQAVQSHAVAEEARKAHQLWLVEALKYSMKTSSAETPTIPLGSAVEAIKANCSDNE  
FTQALTAIIPESLTRGVYSEETLRARFYAVQKLARRVAMIDETRNSLYQYFLSYLQSLLFPPQQ  
LKPPPELCPEDINTFKLLSYASYCIEHGDLELAAKFVNQLKGESRRVAQDWLKEARMTLETQKIVE  
ILTAYASAVGIGTTQVQPE

>HP28  
PPADSLKYDTPVLSRNTEKRSRKARLLKVSPQQPGPSGSAPQPPKTKLPSTPCVPDPTKQAEI  
LNAILPPREWVEDTQLWIQQVSTPSTRMDVVHLQEQLDLKLQQRQARETGICPVRRELYSQCFDE  
LTREVTINCAERGLLLLVRDEIRMTIAAYQTLYESVAFGMRKALQAEQGKSDMERKIAELETEK  
RDLERQVNEQKAKCEATEKRESERRQVEEKHNEBIQFLKRTNQQLKAQLEGIIAPKK

>HSPC232  
RRRADGCIYGVSRRARVVAYRRDEMWSGRYERYERIPRERAPPRSHPSDESGYRWRDDHSASRQP  
EYRDMRDGFRRKSFYSSHARERSPYKRDNTFFRESPVGRKDPSPHSRGSSVSSRSYSPERSKSYS  
FHQSQHRNKERPVQSLKTSRDTSPSSGSAVSSSKVLDKPSRLTEKELAEEASKWAAEKLEKSDESN  
LPEISEYEAGSTAPLFTDQPEEPESNTTHGIELFEDSQLTTRSKAIASKTKEIEQVYRQDCETFGM  
VVKMLIEKDPSEKSIQFALRQLHEIGERCVEELKFIAEYDTSTQDFGEFF

Figure 6 (continued)

>HYPA  
 GRRSSLSPTMRPGTGAERGGMMGHPGMHYAPMGHPMGQRANPPVPHGMPQMMPPMGGPPMG  
 QMPGMMSSVMPGMMMSHMSQASMQPALPPGVNSMDVAAGTASGAKSMWTEHKSPDGRTYYNTETK  
 QSTWEKPDDLKTPAEQLSKCPWKEYKSDSGPKYYNSQTKESRWAKPKELEDLEGYQNTIVAGSL  
 ITKSNLHAMIKAEESSKQEECTTISTAPVTTEIPTTMSTMAAAEEAAAVVAAAAAAAANA  
 NASTSASNTVSGTVPVPEPEVTSIVATVVDNENTVTISTEEQAQLTSTPAIQDQSVEVSSNTGEE  
 TSKQETVADFTPKEEEEESQPAKTYTWNTKEEAKQAFKELLKEKRVPNASWEQAMKMIINDPRY  
 SALAKLSEKKQAFNAYKVQAKKEKKKKKK  
>HZPH  
 HARFAEAECCLAESHQHLSKESLAGNKPANAVLHKVILNQLEELLSDMKADVTRLPATLSRIPPIAAR  
 LQMSERSILSRSLASKGTEPHPTPAYPPGPYATPPGYGAFAFSAAPVGALAAAGANYSQMPAGSFITA  
 ATNGPPVLVKKEKEMVGALVSDGLDRKEPRAGEVICIDD  
>IKAP  
 LKEGSPLEDLALLEALSEVVQNTENLKDEVYHILKVLFLFEFDEQGRELOKAFEDTLQLMERSLPE  
 IWTLTYQQNSATPVLPVNSTANSIMASYQQQKTSVPVLDALFIPPKINRRTQWKLSLLD  
>IMPD2  
 DFLILPGYIDFTADQVDLTSALTKitLkTlVlSSPMdTvtEAGMAIAmALTGGIGFIHHNCTPEF  
 QANEVRKVKKYEQGFITDPVVLSPKDRVRDVFEAKARHGFCGIPITDTGRMGSRLVGIISSRDIDF  
 LKEEEHDCFLEEIMTKREDLvvApAGITLKEANEILQRSKKGKLPiVNEDELVAIIARTDLKKNR  
 DYPLASKDAKKQLLCGAAIGTHEDDKYRLDLLAQAGVDVvVLDSSQGNSIFQINMIKYIKDKYPNL  
 QVIGGNVVTAAQAKNLIAGVDALRVGMGGSICITQEVLAAGRQATAVYKVSEYARRFGVPVIA  
 DGGIQNVGHIAKALALGASTVMMCSLLAATTEAPGEYFFSDGIRLKKYRGMGSLAMDKHLSQQNR  
 YFSEADKIKVAQGVSGAVQDKGSIHKFVPYLIAGIQHSCQDIGAKSLTQVRAMMYSGELKFEKRTS  
 SAQVEGGVHSLHSYEKRLF  
>KPNA2  
 AWALTNIASGTSEQTKAVVDGGAIPAFISLLASPHAHISEQAVWALGNIAGDGSVFRDLVICKYGA  
 DPLLALLAVPDMSSLACGYLRNLTWTLSNLCRKNPAPPIDAVEQILPTLVRLHHDDPEVLA  
 WAI SYLTDGPNERIGMVVKTGVPQLVKLLGASELPiVTPALRAIGNIVGTDEQTQVVIDAGALA  
 VFPSSLTNPKTNIQKEATWTMSNITAGRQDQIQQVWNHGLVPFLVSVLSKADFKTQKEAVWAVTN  
 TSCGTVEQIVYLVHCGIEPLMNLLTAKDTKIIILVILDAISNIFQAAEKLGETEKSIMIECGGL  
 DKIEALQNHENESVYKASLSLIEKYFSVEEEEDQNVVPETTSEGYTFQVQDGAPGTFNF  
>KPNB1  
 LAAVGLVGDLCRALQSNIIIPFCDEVMQLLLENGNENVHRSPVKPQILSVPGDIALAIGGEFKYLE  
 VVLNTLQQASQAQVDKSDYDMVDYLNELRESCLEAYTGIVQGLKGDQENVHPDVMLVQPRVEFILS  
 FIDHIAGDEDHTDGVVACAAGLIGDLCTAFGKDVLKLVEARPMIHELLTEGRRSKTNKAKTLATWA  
 TKELRKLKNQA  
>KU70  
 KTRTFNTSTGGLLLPSDTKRSQIYGSRQIILEKEETEELKRFDDPGLMLMGFKPLVLLKKHHYLRP  
 SLFVYPEESLVIIGSSTLFSALLIKCLEKEVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPP  
 GFQLVFLPFADDKRKMPTEKIMATPEQVGKMKAIVEKLRFTYRSDFSVPVLQOHFRNLEALALD  
 LMEPEQAVELTLPKVEAMNKRLGSLVDEFKELVYPPDYNPEGKVTKRKHDNEGSGSKRPKVEYSEE  
 ELKTHISKGTLGKFTVPMLEACRAYGLKSGLKKQELLEALTKHFD  
>LUC7B1  
 VDAVAVDAAAASAKAEKVHELNEKIGKLLAKAEQLGAEGNVDESQKILMEVEKVRACKKEAEEYR  
 NSMPASSFQQQKLRVCCEVCSAYLGLHDNDRRLADHF GGKLHLGFIQIREKLDQLRKTVAEKQEKR  
 QDRLRRREEREREERLSRRSGSRDRRRRSRSRSTSERRKLSRSRSRDRHRRRSRS  
 RSHSRGHRRASRDRSAKYFSSRERASREEWSGRSERGPPDWRLESSNGKMASRRSEEKEAGEI

Figure 6 (continued)

>MAGEH1  
ASFPRTAVSFEPLAGDMPRGRKSRRRNARAAEENRNNRKIQASEASETPMAASVVASTPEDDLSG  
PEEDPSTPEEASTTPEEASSTAQAQKPSVPRSNFQGTKKSSLMSILALIFIMGNSAKEALVWKVLG  
KLGMQPGRQHSIFGDPKKIVTEEFVRGYLIYKPVPRSSPVEFFWGPRAHVESSKLKVMHFVAR  
VRNRCSKDWPCNYDWDSDDDAEVEAILNSGARGYSAP

>MAP11c3  
QRSPFADRCKEVQQIRDQHPSKIPVIERYKGEKQLPVLDKTKFLVPDVHNMSELVKIIIRRLQLN  
PTQAFFL VNQHSMVS VSTPIADIYEQEKD EDFGFLYMVYASQETFGF

>mHAP1  
PKEQVQSGAGDGTGS GDP AAGTPTT QPAVG PAPEPSAEPKPAPA QGTGSGOKSGSRTKTGSFCRSM  
IIGDS DAPWTRYVFQGPYGP RATGL GTKAEGIW KTPAAYIG RRP GVSG PERAAFIRELQEA LCPN  
PPPTKKITTEDDVKVMLYLLEEKERDLNTAARIQGS LVKQNSVLMEE NNKLETMLGSAREEILHLRK  
QVNLRDDLLQLYSDSDDDDDEEDEEDEEEGEEEREQQRDQDQHQHDHPYGA PKPHKAETAHRCPO  
LETLOQKLRLLLEEENDHLREEASHLDNLEDEEQMLILEC VEQPSEASOQMAELSEV LVRLEG YER  
QOEITQLQAEITKLQQRQCQSYGAQTEKLQQLASEKG IHS ESEL RAGSYM QDYGS RPRDRQEDGKS  
HRQRSSMPAGSVTHYG SVPLD ALPS F PETLA EELRTSLRKFI TDPAYFMERRDTHCREGRKKEQR  
AMPPP PAX

>mp53  
VTE TPGPVA PAPAT PWPLSSF VPSQ KTYQ GNYGF HLGFL QSGTAKS VMCT YSPLN KLF CQL A KTC  
PVQLWVSATPPAGSR VRAMAIYKKSQHMTEVVRCPH HERCS DGD GLAPP QHL IRVE GNLYPE YLE  
DROQTFRHSVVV PYEPPEAGSE YTTIHYK YM CNSSCM GMN RR PILT I ITLED SSGN LLGRD SFE VR  
VCACPGDR DR RTTEE ENFRK KEVLCPELPGSAKRAL PTCTSASPPQKKPLDGEYFTLK IRGRKR FE  
MF RELNEALELK DAHATE ESGD RAHSS YLKT KGQ STSRH KTMV KK VGP DSD

>NAG4  
RDRVNEAEKDLQCHAPVR LD LIPPEKPLTSSLAKQEEVEQTPLQEA LNQLMRQLQRKDPSAFFSFP  
VTDFIAPGYS MIIKHPMD FSTMKEKIKNDYQSIEELKDNFKLMCTNAMIY NKPETIYYKA AKKLL  
HSGMKILS QERI QSLK QOSIDF MADLQ KTRK QKD GTDT SQSG E DGGC W QRER E DSG DAE AHAF KSPS  
KENKKDKDM LEDKF KSNN LERE QE QLDR IVK ESGG KLT RRL VNS QCE FERR KPD GTT LGLLHPV  
DPIVGE PGY CPV RLGM TT GRL QSG VNTL QGF KED KR NKV TPV LYLN YGP SSY A PHY D STF AN ISK  
DDS DLI YSTY GE DSD LP SD FS IHE FLAT CQD YPV MADS LLD VLTK GH SRT LQEM EM SLP ED EGH  
TRTL DTAKEME I TEVE PP GR L DS ST QDR LIA L KAV TN FG VP VE VF DSEE AE I FOK KLD ETR LL RE  
LQEAQNERI LSTR PPN MICL LG PSY REM H LA EQ VT NN L KELA QQ VTP GDIV STY GVR KAM GIS IPS  
PVMENN FVD L TED TEEPK KTD VAE CG PG GS

>NEFL  
LSP LSS LSG LPPP PRAGE PPA AT MSSFS YEP YS TS YK RRY VET PRV HISS VR SGY STARS SAY SSY  
SAPVSS SLV RRSY SSS SGSL MPSLEN L DLSQ VAAI SNDLK SIRT QEK AQL QD LND RFAS FIER VH  
ELEQ QNK VLEA ELL VLR QKH SEPSR FRAI LYQE I RD LRLA EADAT NEKO ALQ GERE GLEET LR NLQ  
ARYEE EVL SREDAE GRL MBE ARKA GDA EA ALA ELE KRD SLM DEI SFL KK VHEE EIA LQ A QI QYA  
QIS VEM DVTK PDLS AALKD I RAQ YEK LA KNM QNA EEW FK SRF TVL TES A KNT DVA AKD E VSE  
SRR LLKAKT LEI EAC RGM N EA LEK QL QELED KQ NAD IS AM QD TINK L E N LRT KSE MARY LKEY Q  
DL IN V KMA D LIE I A A YR K LLE GE ETR L SFT SVGS I TSG YS QSS QV FGR SAY GGL QTSS YLM STR SF  
PS YYT SHV QEB QI E VEET IE A A KAE EAKD EPP S EGE A EEE E K D KEE A EEE A EEE A K B E S E E A  
KEEE EGGE GEE GE ET KE A EEE E K VEG AG E E Q A K K D

>p53  
MEPQSDPSVEPLSQETFSDLWKL LPENN VLSPLPSQAMDDMLSPDDIEQWFTEDPGPDEAPRM  
PEA APPVAPAPAAPTAA PAPAPSWPLSSVPSQ KTYQ GSYG FRLGFLHSGTAKS VTC T YSPALNK  
MFCQLAKTCPVQLWV D ST PPPGTRV RAMAIYKQS QHMTEVVRCPH HERCS DGD GLAPP QHL IRVE  
GNL RVE YL D DR NTFR HS VVV PYEPPEVG SDCTT I HY NY MC NSSCM GMN RR PILT I ITLED SSGN L  
LGRNSF E V RV CACP GDR RTTEE ENLR KK GE PHBL PPG STK RAL PNNTTSS SPQPK KPLDGEYFTL  
QIRGRER FEM FRELNE ALELK DAQ AGK EPG GS RAHSS H LK S K K Q STSR H K L MF K TEG P DSD

Figure 6 (continued)

>PFN2  
APRRPRCSAKGSKAGWQSVDNLCDGCCQAAIVGYCDAKYVWAATAGGVFQSITPIEIDMIVG  
KDREGFFTNGLTLGAKKCSVIRDSLVDGDCTMDIRTKSQGGEPTYNAVGRAGRVLVFVMGKEGV  
HGGGLNKAYSMAKYLRDGF  
>PIASy (bait)  
LVEAKNMVMSFRVSDLQMLLGFGVGRSKSGLKHELVTRALQLVQFDSCP ELFKKIKELYETRYAKKN  
SEPAPQPHRPLDPLTMHSTYDRAGAVPRTPLAGPNIDYPVLYGKYLNGLGRLPAKTLKPEVRLVKL  
PFFNMLDELLKPTELVPQNNEKLQESP C IFALT P R Q V E L I R N S R E L Q P G V K A V Q V V L R I C Y S D T S C  
PQEDQYPPNIAVKVNHSYCSVPGYYPSNKGVEPKRPCRPI NLTHL MYLSSATNRITVTWGNYGKS  
YSVALYLVRLQTSSELLQRLKTIGVKHPELCKALVKEKLRLDPDSEIATTGVRVSLICPLVKMRLS  
VPCRAETCAHLQCFDAVFYLQMNEKKPTWMCPCDKPAPYDQLIIDGLLSKILSECEADEIEYLV  
DGSWCPIRAEKERSCSPQGAILVLG PSDANGLLPAPS VNGSGALGSTGGGPVGSMENGKPGADVV  
DLTLDSSSSSEDEEEEEEEDEDEEGPRPKRRCPFQKGLVPAC  
>PIASy (prey)  
LVEAKNMVMSFRVSDLQMLLGFGVGRSKSGLKHELVTRALQLVQFDSCP ELFKKIKELYETRYAKKN  
SEPAPQPHRPLDPLTMHSTYDRAGAVPRTPLAGPNIDYPVLYGKYLNGLGRLPAKTLKPEVRLVKL  
PFFNMLDELLKPTELVPQNNEKLQESP C IFALT P R Q V E L I R N S R E L Q P G V K A V Q V V L R I C Y S D T S C  
PQEDQYPPNIAVKVNHSYCSVPGYYPSNKGVEPKRPCRPI NLTHL MYLSSATNRITVTWGNYGKS  
YSVALYLVRLQTSSELLQRLKTIGVKHPELCKALVKEKLRLDPDSEIATTGVRVSLICPLVKMRLS  
VPCRAETCAHLQCFDAVFYLQMNEKKPTWMCPCDKPAPYDQLIIDGLLSKILSECEADEIEYLV  
DGSWCPIRAEKERSCSPQGAILVLG PSDANGLLPAPS VNGSGALGSTGGGPVGSMENGKPGADVV  
DLTLDSSSSSEDEEEEEEEDEDEEGPRPKRRCPFQKGLVPAC .  
>PLIP  
GEIIEGCR LPV LRRNQDNEDEWPLAEI LS V K D I S G R K L F Y V H Y I D F N K R L D E W V T H E R L D L K K I Q F  
PKKEAKTPTKNGLPGSRPGSPEREVKRKVEVVSPATPVPSETAPASVFPQNGAARRAVAAQPGKR  
KSNCLGTDEDSQDSSDGIPSAPRMGTGSLVSDRSHDITVTRMKNIECIELGRHRLKPWYFSPYPQEL  
TTLPVLYLCEFCLKYGRSLKCLQRHLTCKCDLRHPPGNEIYRKGTISFFEIDGRKNKSYSQNLCLLA  
KCFLDHKTLYYDTDPFLFYVMTEYDCKGFHIVGYFSKEESTEDYNVACILTLPYQRRGYGKLLI  
EFSYELSKVEGKTGTPEKPLSDLGLLSYRSYWSQTILEILMGLKSESGERPQITINESEITSIKK  
EDVISTLQYLNLINYYKGQYILTLSEDIVDGH ERAMLKRLLRIDS KCLHFTP KDW SKRGK  
>PTN  
LSQRQDQVPRLPVQKS RQES PRAEENPKWREGKETSESSVQKAGR AAAA QAGAA ASRV PG LSGSN  
LAPCNKGRLSAREDVNSNSKMQAQYQQQRKFAAFLAFIFILA AVDTAEAGKKEKPEKKVKKSDC  
GEWQWSVCVPTSGDCGLGTREGTRTGAECQTMKTQRCKIPCNWKKQFGAECKYQFQAWGECDLNT  
ALKTRTGSLKRALHN AECQKTVTISKPCGKLTKPKPQAESKKKKKEGKKQEKMLD  
>PTPK  
SNYINAALMDSYRQPAAFIVTOYPLPNTVKDFWRLVYDYGCTSIVMLNEVDLSQGCPQYWPEEGML  
RYGPIQVEC M S C S M D C D V I N R I F R I C N L T R P Q E G Y L M V Q Q F Q Y L G W A S H R E V P G S K R S F L K L I L Q V  
E K W Q E E C E E G E G R T I I H C L N G G R S G M F C A I G I V V E M V K R Q N V V D F H A V K T L R N S K P N M V E A P E Q  
Y R F C Y D V A L E Y L E S S  
>SETBD1  
KASTSGLGIKDEGDIKOAKKEDTDDRNKMSVVT ESSRNYGYNPSPVKPEGLRRPPSKTSMHQS RRL  
MASAQSNPDDVLTLSSTESEGESEGT SRKPTAGQTSATAVDSDDIQTISSGSEGDDFEDKKNMTGP  
MKRQAVKSTRGFALKSTHGIAIKSTNMA SVDKGESAPVRKNT RQFYDGE ESCYI IDAKLEGNLGR  
YLNHSCSPNLFVQNFVDT HDLRF PW AFFASKRIRAGTEL TD NYEVGSVEG KELLCCC GAI EC  
RGRLL .

Figure 6 (continued)

>SH3GL3  
 VAGLKKQFHKAQLFSEKISGAEGTKLDDEFIDLMERKIDVTNKVVAEILSKTEYLQPNPAYRAKL  
 GMLNTVSKIRGQVKTTGYPQTEGGLGDCMLKYKGELGEDSTFGNALIEVGESMKLMAEVKDSLIDIN  
 VKQTFIDPLQLLQDKDLKEIGHLKLEGRRLDYDYKKRUGKIPDEEVVRQAVEKFEESSKELAERS  
 MFNFLENDVEQVSQALAVFIEAALDYHRQSTEILQELQSKLQMRISAASSVPRREYKPRPVKRSSSE  
 LNGVETTSVVKTGNSNIPMDQPCRGLYDFEPENQGELGPKEGDIITLTNQIDENWYEGMIHGESG  
 FFPINYVEVIVPLPQ  
>SUMO-2  
 RPRAQLRRESGGAESVTRPLRAASPAPPRAARAAMSEEKPKEGVKTENDHINLKVAGQDGSVVQF  
 KIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQUEMEDEDTIDVFQQQTGGVPESS  
 LAGHSF  
>SUMO-3  
 PSSTAAASFFCRSWCCLCARLVRTWYLFCCEAAAEEETPALAMADEKPKEGVKTENNDHINLKVAGQD  
 GSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQUEMEDEDTIDVFQQQTG  
 GVY  
>TAL1  
 SSPVKRQRMEASLDQLKQFTTVVADTGDFHAIDEYKPQDATTNPSLILAAAQMPAYQELVEEAIAY  
 GRKLGGSQEDQIKNAIDKLFVLFGAEILKKIPGRVSTEVDARLSFDKDMVARARRLIELYKEAGI  
 SKDRILIKLSSTWEGIQAGKELEEQHGIHCNMILLFSFAQAVACAEAGVTLI SPFVGRILDWHVAN  
 TDKKSYEPELEDPGVKSVTKIYNYKKFSYKTIVMGASFRNTGEIKALAGCDFLTISPKLLGELLQD  
 NAKLVPVLSAKAAQASDLEKIHLDKESFRWLHNEDQMAVEKLSDGIRKFAADAVKLERMLTERMFN  
 AENGK  
>TCPG  
 QTDIEITREEDFTRILQMEEEYIQQLCEDI IQLKPDVVITEKGISDLAQHYLMRANITAIRRVRKT  
 DNNRIARACGARIVSRRPEELREDDVGTGAGLLEIKKIGDEYFTFITDCKDPKACTILLRGASKEIL  
 SEVERNLQDAMQVCRNVLLDPQLVPGGGASEMAVAHALTEKSKAMTGVEQWPYRAVAQALEVIPRT  
 LIQNCASTIRLLSLRAKHTQENCETWGVNGETGTLVDMKELGIWEPLAVKLQTYKTAVETAVIL  
 LRIDDIVSGHKKGDDQSRRQGGAPDAGQE  
>VIM  
 SPRQRSSRAPTTTHRALVRLFSCSQSAPPPPPRSPSSAAMSTRVSSSYRRMFGGPGTASRPS  
 SSRSYVTTSTRTYSLGSALRPSTSRSILYASSPGVYATRSSAVRLRSSVPGVLLQDSVDFSLADA  
 INTEFKNTRTNEKVELQELNDRFANYIDKVRFLEQQNKLILLEQQLKGQGKSRLGDLYEEEMREL  
 RRQVDQLTNDKARVEVERDNLAEIDIMRLREKLQEEMLQREEAENTLQSFQDVDNASLARLDLERK  
 VESLQEEIAFLKKLHEEEIQELOAQIQEQQHVQIDDVDVKPDLTAAALRDVRQQYESVAAKNLQBAEE  
 WYKSKFADLSEAANRNNDALRQAKQESTEYRRQVQSLTCEVDALKGTNESLERQMREMEENFAVEA  
 ANYQDTIGRLQDEI QNMKEEMARHLREYQDLLNVKMALDIEIATYRKLEGEESRISLPLPNFSSL  
 NLRETNLDSLPLVDTHSKRTLLIKTVETRDGQVINETSQHDDLE  
>VIMC  
 QEEMLQREEEAENTLQSFQDVDNASLARLDLERKVESLQEEIAFLKKLHEEEIQELOAQIQEQQHVQ  
 IDDVDVKPDLTAAALRDVRQQYESVAAKNLQEAEEWYKSKFADLSEAANRNNDALRQAKQESTEYRR  
 QVQSLTCEVDALKGTNESLERQMREMEENFAVEAANYQDTIGRLQDEI QNMKEEMARHLREYQDLI  
 NVKMALDIEIATYRKLEGEESRISLPLPNFSSLNLRETNLDSLPLVDTHSKRTLLIKTVETRDGQ  
 VINETSQHDDLE

Figure 6 (continued)

>ZHX1  
EQTINDLTFDGSFVKEENAEQAESTEVSSSGISISKTPIMKMMKNKVENKRIAVHHNSVEDVPEEK  
ENEIKPDREEIVENPSSSASESNTSTSIVNRIPHSTASTVVTPAAVLPGLAQVITAVSAQQNSNLI  
PKVLIPVNSIPTYNAALDNPLLNTYNKF PYPTMSEITVLSAQAKYTEEQIKIWFSAQRLKHGV  
WTPEEEVEARRKQFNGTVHTVPQTITVIPTHISTGSNGLPSILQTCQIVGQPGLVLTQVAGTNL  
VTAPIALTAVGVPSQNNIQKSQVPAAQPTAETKPATAAVPTSQSVDKETALVNPDSFGIRAKKT  
QLAELKVSYLNQFPHDSEIIIRLMKITGLTKGEIKKWFSDTRYNQRNSKSNQCLHLNNDSSTTII  
DSSDETTESPVGTAAQPKQSWNPFPDFTPQKFKEKTAEQLRVLQASFLNSVLTDEELNRLRAQTK  
LTRREIDAWFTTEKKKSALKEEKMEIDESNAGSSKEEAGETSPADESCAPKGSTGKICKKTPEQL  
HMLKSAFVRTQWPSPPEEYDKLAKESGLARTDIVSWFGDTRYAWKNGNLKWYYYYQSANSSSMNGLS  
SLRKGRGRGPKGGRGRPRGRPRGSKRINNWDRGPSLIKFKTGTAIKDYYLKHKFLNEQDLDELV  
NKSHMGYEQVREWFAERQRSELGIELFEENEEDEVIDDQEEDEEETDDSDTWEPPRHVKRKLSK  
SDD  
>ZNF33B  
CYECGKTFCLKSDLTIHQRTHTGEKPFACPECGKFFSHKSTLSQHYRTHTGEKPYECHECGKIFYN  
KSYLTKHNRTHTGEKPYECNECGKTFQKSQLTQHQRIHIGEKPYECNECGKAFCHKSALIVHQRT  
HTQEKPYKCNECGKSFCKVSGLILHERKHTGEKPYECNECGKSFHKSSLTVHYRAHTGEKSCQCN  
ECGKIFYRKSDLAKHQRSHTGEKPYECNTCRKTFSQKSNLIVHQRTIGEKPYE

Figure 6 (continued)

&gt;ALEX2

GCCGAATCAGTAGTTGGGCTGCAATGGCTCTGCAATAGCACCACTCCCGGGGTGACAGAGGCC  
 CTTGGGCTGCAGAAGCCCCCTGCAATGGCAGGGCTCCCAAAGTGGCAGAAGCTCCAGAGAACGG  
 GAGACTTCCAGGGCAGCGGTGCCTCTGGACAGTGGTGCTACCGAAGCGGCAGCACCACTGAG  
 GTGACCGAGGGTCTGGGTAGCAGCACCTACCAAGGTAGCTGAAGCTCCGGGTGGCATGCC  
 ACCGAGGCAGCTGAGGCTCTGTGCCGCAACGCCACTGGGCTGCAGCACCTACTGGGCTGCA  
 GAGTCTCCTGAACTTCTGGTCCCCTAGAACACAGCGGTGTTCTGGAACATCAGCTGCCAAGAAA  
 GCAACCCCTGGGCTCACACTGGGCTATAACGAAAGCCACATCAGCAGCTGGAGCGGTACCCAAA  
 GGTGGAGGCAAGGGTGTAAACCAAGGTCCCGAATGGGCAAGGGCAAGGGCAAGAAAAGCAAAGTT  
 GAAGTAGACGAACCTGGGATGGCTTCCGTCTGGAGATGGGCTGCAGCACCTGCTGCAGCCTCT  
 GCTAATGGCGACAGGCTTCTGGCAGAGGTCCCTGATTCTGAGGAAGGGAGTCCGGGTGGACT  
 GACACAGAGTCAGATTCAAGACTCTGAGCCCCAGCCAGGGAGGGGAAGAAGACCCGTT  
 GCCATGCAAGGCCCTTCTTATGAAATTGATGAGATTCTGGGTGTCCCGATCTCAGGAAG  
 GTCTTGCCTGCTTCAGAAATCTGATGATCCTTCATCCAACAGGTAGCTTCTCACTCTGAGC  
 AACAAATGCCAATTATTCAATGCAATCAAGAGACAATCCGCAAATTGGGAGGCCCTCCAATTATTGCA  
 AACATGATCAACAAAATGATCCACACATTAAGGAAAAGCCTTAATGGCCATGAATAACCTGAGT  
 GAGAATTATGAAAATCAGGGCCGGCTCAGGTGTACATGAATAAAAGTGTGATGGATGATCATGGCC  
 TCTAACCTGAACCTCAGCAGTCAAGTAGTTGGACTAAAAATTCTAACAAACATGACTATTACTAAAT  
 GACTACCAACACCTGCTGCAATTCCATTGCAAACCTTTCCGTTGCTATCTCAGGGAGGTGGA  
 AAAATCAAGGTGAGATTGAAAATCTTCTGAAATTGCTGAAAATCCAGATATGTTGAAGAAA  
 CTTCTCAGTACCCAAAGTGCCAGCATCTTGTACCTCTATAATTCTTACGTGGAATCAGAAATC  
 CTTATTAAATGCCCTTACTCTATTGAGATTATCTATGACAATCTCAGAGCAGAAAGTGTAACTAT  
 AGAGAAATTCAATAAAAGGTTCCCTTTTACTTATGCACATCTGGAGTGTGTTAGAAAATT  
 AGAGCCTAGCAAATCACCAGACCTCTTAGTGAAGTATAAAACTAGTGAACAAATTCA

&gt;APP1

GAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCTCCGCAGGCTGAA  
 GAGGAAGAGGAAACGGTCCCACCCCCAACGCTCCATACACTTGCAGTGGTCGGCAAAGTCACTCCC  
 ACCCGAGGCCACAGACGGTGTGGATATTACTTGGCATGCCTGGGAAATCAGTGAGCACGAG  
 GGGTCCCTGAGGGCCAAGATGGACCTGGAGGAGCGTAGGATGCGCAGATTAATGAGGTGATGCGT  
 GAATGGGCCATGGCAGACAACCAAGTCCAAGAACCTGCCCTAAAGCCGACAGACAGGCCCTGAATGAG  
 CACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCAGGTGCTGGTGAGCGACAGCGCCTGGTGGAA  
 ACCCACGCCACCCCGGTATCGCCCTTATCAACGACCAGCGCCGGCTGCCCTGGAGGGCTTCTG  
 GCAGCCCTGCAGGAGATCCGCTCAGGCGGAGCGTCTGGCCCTGCGCGCTACCTGCGT  
 GCGGAGCAGAAGGAACAGAGGCACACGCTGCCACTACCAAGCATGTGGCCGCCGTGGATCCCGAG  
 AAGGCACAGCAGATGCGCTTCAGGTGCATACCCACCTCAAGTGATTGAGGAGAGGGTGAATCAG  
 AGCCTGGGCTGCTGACCAGAACCCCCACCTGGCTCAGGAGCTGGGCCCCAAATCCAGGAACCTC  
 CTCCACTCTGAACACCTGGTCCCAGTGAATTGAAAGGCCCTGCCCTGGGGCAGCAGCGAGGAC  
 AACGGTGGGCTGCAGCCTCCAGATTCCAAGGATGACACCCCCATGACCCCTCCAAAAGGGTCCACA  
 GAACAAAGATGCTGCATCCCCTGAGAAAGAGATGAACCCGCTGAAACAGTATGAGCGAAAGGTG  
 AATGCGTCTGTTCCAGGGTTTCCCTTCACTCATCGGAGATTAGAGGGATGAGCTGGCACCAAG  
 CTGGGACAGGGTGTCCGTGAGGCTGTGCGGGTCTGC

Figure 6 (continued)

&gt;BAIP1

CGGCCGGACGAAGATGGCGACGCCATGTACTTGGAGCACTATCTGGACAGTATCGAGAACCTT  
 CCCTGCAGACTTCAGAGGAACCTCCAGCTGATGCGAGAGCTGGACCAGAGGACGAAAGATAAGAAA  
 GCAGAGATTGACATCCTGGCTGCAGAGTACATCTCACGGTAAGACGCTGTCTCCAGACCAGCGC  
 GTGGAGCGCCTGCAGAAGATCCAGAACCCCTACAGCAAGTGAAGGAATACAGTGACGACAAAGTG  
 CAGCTGGCATGCGACACCTACGAGATGGTGGATAAACACATCGAAGGCTTGATGCAGACCTGGCG  
 CGCTTGAAGCAGATCTGAAGGACAAGATGGAGGGCAGCGATTTGAAAGCTCCGGAGGGCGAGGG  
 TTAAGGAGGGCTCAGAAAGAAAAAGAGGGTCCCAGGGCGAGGCAGGAGGACATCAGAG  
 GAAGACACACCAAAGAAAAAGAACGACAAAGGAGGGTCTGAGTTACTGACACCATCCTGTCCGTG  
 CACCCCTCTGATGTGCTGGACATGCCGTGGACCCAAACGAACCCACGTACTGCCTGTGCCACCAAG  
 GTCTCCTATGGGAGATGATTGGCTGTGACAATCAGACTGTCCAATTGAGTGGTTCACTTGCC  
 TCGTGGACCTTACACGAAACCCAAAGGAAAATGG

&gt;BAIP2

AGCCAGCAGGCCAGCGTGACCATGCACGATGTGGACGCCAGTCCTCGAGGTGTTGGTCGACTAC  
 TGCTACACGGGTCGTTCTCAGTGAGGCCAATGTGAGCGCCTGTACGCGCCCTCCGACATG  
 CTACAGCTGAAATATGTGCGGAAGCCTGTGCTCCTCTTAGCCCCGACGCTTGACCTGACCAAC  
 TGCAACGCCATCCTCAAGTTGCAAGACGCCCTCGACCATCACAAGCTTGATCTCAGGCCAGTCC  
 TACATAGCTCACAACCTCAAGCAGCTCAGCCGAATGGGTTCAATTCTGGGAGGAGACTCTAGCAGAT  
 CTAACCCCTGGCCAGCTGCTGGCTCCTACGCCCTGGATAGTCTGGACATAGAGAGTGAGCGGACT  
 GTATGCCATGTAGCTGTGCACTGGCTGGAGGCTGTCGCAAAGAGCGGGGCTCCAGTGTGCAGAA  
 GTCTTCAAGTGCCTGCGCTGGATGCACTTCACTGAAGAAGATCAGGACTACTTAGAAGGGCTGCTG  
 ACCAAGCCCCTCGTGAAGAAGTACTGCCCTGGACGTTATTGAAGGGGCCCTGCAGATGCCATGGT  
 GACCTGTTGACAAAGTCTGCTGGCTCAGTGCACACAGCAGCAGCAGTAGCAGCAGCAACTCT  
 CTTGTATCTGCAGCAGAAAATCCACCCAGAGACTGGGTATGTGCAAGGGAGATGGTATCTC  
 TTTGGACATCCTAGAGATCCCTTCTGCTATGACCTTACTCGGGGACATTACACAAATGCCA  
 TCCCTTGGACAGCTTGCTCACACTAAGACTGTCACCTCTCAGCTGCTGTGTGCCCCAGAC  
 CATGACATCTATCTAGCTGCTCAGGCCAGGAAAGACCTCTGGGTGATAAACCAAGCTCAGAAATGT  
 TGGCAGCAACTTGCAAGATGCCCTGCTGTCGAGGGCATGGATGTCATATCTCAATGGCTAC  
 ATCTACATTGGGGGACGAGACCTTAACTACTGAGTTAAGTGAAGGAAGTGGAAATGCTACAGT  
 GTTCAGAGAAACCAAGTGGCATTGGCTCCTGCTCCTATTCCCTCTATTCCCTTGAACTCATA  
 GTGGTTCAAGAAACTATCTTATGCTGTCACACTAAGCGCAGTGTGCTTGTGCTATGATCCTAGCCACAAT  
 ATGTGGCTGAAGTGTGCTCTTAAACGTACTGACTTTCAGGAAGCAGTGTGCTTCAATGATGAA  
 ATCTATTGTATCTGTGACATCCAGTCAGAAGGCTTACACCCAGCTAGGGAGAATGGAGGCGG  
 ATTAGTAATATTCCATTGGATTCAAGAGACCCACAACCTACCAAGATTGTCATGACCAAAAGTTG  
 CTTCTCATCACTTCAAAACCCACAATGGAAAAAGAACCGAGTGACAGTGTATGAGTATGATACT  
 AGGGAAAGATCAGTGGATTAAATATAGGTACCATGTTAGGCCTTTGCAGTTGACTCTGGCTTATT  
 TGCCTTGTGCTCGTGTGTTATCCTTCCCTGCGCTTGAACCTGGTCAGAGTTTATTACTGAGGAAGAT  
 GATGCACGGAGTGAAGTCTAGTACTGAATGGGAAGTGAAGTGGACTCTGAGTCA  
 GGAAGTTCAAGTTCTTTCAAGATGATGAAGTCTGGGTGCAAGTAGCACCTCAGCGAAATGCACAG  
 GATCAGCAGGGTTCTTG

Figure 6 (continued)

>BAIP3  
 GGACACAATGCCCCAGAAAAGTAACAGCCGTCAATTATGCTAGAAAAGGAAGTGTCTCCAGAGC  
 ATAGAGAAAATAAGTTCTCTGTGATGCAACAACTGTTACTTCACAACAGTGTGTTTCAGAGAC  
 CAAGAACCAAAGATCCATAATGAGATGGCATCACACATCAGATAAAGGTGCCAAGGAAGAAATGAC  
 AAGAAAGATTCTCAAGGAAGAAGTAATAAGGCATTACATCTGAAGAGTGATGCTGAATTAAAAAG  
 ATATTGCCCCCTACTAAGGATTGAGAGTGTCCTTACTCGAATTCTGACCATTGACCTCTGGA  
 GAAGGTTTCGATTCTTAGCAGTTGGTAAAGAGTGGTACTTACAAGAGACAGAGTTATGGTG  
 AAGGAAGGAGAGAGAAAACAGCAGAATTGATAAGAAAAGCAAAACTAATAAGAAGATG  
 GATCACATAAAGAAGAGAAAACAGAGAATGCTTATAACGCAATCATAATGGGAAGCTAATGTC  
 ACCGGTCTCCAACTCTAAGCAGTATTTCACCAACTTCAGATGTGTCACAACATAACATTCTCACG  
 AGTCACAGCAAAACCAAGACAAGAAAAGAGAACTGAGATGGAATACTATACCCATGAGAAGCAAGAG  
 AAAGGCACTTGAATTCAAATGCACTTATGAACAAAGTCATTCTCAATAAAAATTATACCGAA  
 GATATTCTCCAGTGACACCACCGGAGTTAGAAGAAACCATTGAGATGAAAAAATAAGAAGACTT  
 AAGCAGGTGCTGAGAGAGAAGCAGCTTGAAGAAATGCGTAAGAAGATGCCACCAAAA  
 >BARD1  
 TTGGCCGGTTTCGAGTCGCTGACCTGCAGCTTCCCTGTGGTTCCCGAGGCCCTCTGCTTCCCGC  
 TCTCCGGAGGAGCCTTCATCCGAAGGCGGGACGATGCCGATAATCGCAGCCGAGGAACCAGGCA  
 CCGAGGATCCGCTCCGGAACGAGCCTCGTTCCCGCTGCCATGGAACCGGATGGTCGCGGTGCC  
 TGGGCCCCACAGTCGCGCCGCGCTGACCGCCTGGAGAAGCTGCTGCGCTGCGTTGACTAAC  
 ATTCTGAGAGAGCCTGTGTGTTAGGAGGATGTGAGCACATCTCTGTAGTAATTGTGTAAGTGAC  
 TGCATTGGAACTGGATGTCCAGTGTGTTACACCCGGCTGGATACAAGACTGAAAGATAAAAGA  
 CAACTGGACAGCATGATTCAACTTGTAGTAAGCTTCGAAATTGCTACATGACAATGAGCCGTCA  
 GATTGAAAGAAGATAAAACCTAGGAAAAGTTGTTAATGATGCAAGGAAACAAGAAGAAATTCAATT  
 AAAATGCGTTAGCCCTCGAAGTAAGAAAGTCAGATATGTTGAGTAAAGCTTCAGTGCAAAC  
 CAGCCTGCAATAAAAAGATGCAAGTGCTCAGCAAGACTCATATGAAATTGTTCCCAAGTCCT  
 CCTGCAGATGTTCTGAGAGGGCTAAAAGGCTCTGCAAGATCTGAAAAAGCAAAAAAGAAA  
 ACTTTAGCTGAAATCAACAAAAATGAAATTAGAGGCAAGGAGATGGTAATTGACTCC  
 AAAGAGGAATCTAAGCAAAAGCTGGATCCTCTGTAGCCAACCATCTGTTATCTCAGTCCTCAG  
 ATAAATGGTAAATAGACTTACAGCAAGTGGCTCTTGACAGAACTGAAATGTTGGAGTTA  
 ACTGAAGTCTCTTACCATGGCTGAGCAAATAGAGTCTCAGACACTAAGAGCAGGAATGAAGTA  
 GTGACTCCTGAGAAGGTCTGCAAAAATATCTTACATCTAAGAAATCTTGCCATTAGAAAATAAT  
 GGAAAACGTGGCCATCACAATAGACTTCCAGTCCCATTCTAAGAGATGTAGAACCGAGATTCTG  
 AGCACCAGTGGAGATTGTTAGCAACGGTGCCCTCAGAAAATATACCATGCTGAATGTTCT  
 TCACCACTTCATGCAAACGTAAGTTGGTGGTACATCAGGGAGCAAAACAGTAACATGTCGATG  
 AATTCAATTAGTCTTCACCAAGGTACACCACCTTACAT  
 >CA150  
 CAACAATTCTGGGCCCTGAAAGATACTTGTGTTGGCCCTGCTGTCTATTAGCCAAGCACCC  
 ACAACACAAGATCAGACCCCAAGTTCTGCTGTTCACTTCACTGCTACAGTTAGTGTTCACCT  
 CCTGCTCCTACAGCCACACCTGTGCAACCGTCTCCAGCCGACCCCTCAGACGTTACCTCCTGCT  
 GTTCCTCATTCTGATACCTCAGCCAACACAGCAATACCTGCTTTCCACCGAGTAATGGTACCTCCG  
 TTCGTGTTCCCTTGGCATGCCAATTCACTTCCAGGTGTATTGCCAGGAATGGCCCTCCT  
 ATCGTACCCATGATAACATCCCAGGTTGCTATTGCACTGCTACCTTAGCTGGAGCAACA  
 GCAGTTCTGAATGGACTGAATATAAACAGCAGATGGGAAGACATATTATAATAATAGAACA  
 TTAGAATCAACCTGGAAAAACCCCAAGAACTAAAGGAAAAGAAAAGTTAGAAGAGAAGATTAA  
 GAGCCAATTAAAGAACCCCTCTGAAGAGCCTCTGCCAATGGAGAGCAGGAGGAGGAGTCTAAAGAA  
 GAGCCTATAAAGGAGATAAAGGAGGAGGCCAAGAAGAGGAGATGACTGAAGAAGAAAAGGCTGCC  
 CAGAAGGCAAAGCCAGTTGCTACTGCTCCTATTCTGGTACTCCATGGTGTGCGTTGGACTGGT  
 GATGAGCGGGTCTTCTTTATAATCCACACTGCTCTTCTATGTGGGACCGACCTGATGATCTG  
 ATTGGCAGGGCAGATGTTGACAAAATTATTAGGAGCCCTCATAAAAAGGAATGGAGGAATTG  
 AAGAAACTAAGGCACCCAACTCCGACAAATGCTGTCGATCCAAAAGTGGCAATTCTATGAGTGCA  
 ATTAAAGAGGAACAAGAATTAAATGGAAGAAATTAAATGAAGATGAGCCTGTTAAAGCAAAAAACGG  
 AAG

Figure 6 (continued)

>CGI-125  
TTCGACCGCTCCCGCGGAACCTCGCCCGCGTCTGGGCTTTGCTCTGTCAAGGCTGGTGGCGTT  
TTGGTGTCTCGTTATGCCGCTGCTCGCTATGGAGACAGATGATGCTGGAAATCGACTT  
CGGTTCACTGGAGTTGAATTGTGCAATGTTAGCCAACCCAAATTACCTTAATTCTTCTGCC  
CAAAGAGGTTACTTCAAACAGACAAAGCTTTGTTAATTATCTAAATACCTGCTTACTGGAAAGAC  
CCAGAATATGCCAAGTATCTAAAGTACCCCTCAGTGTACACATGTTAGAGCTGCTCCAATATGAA  
CACTCCGAAAGGAGCTGGTGAATGCTCAGTGTGCGAAATTATTGATGAACAGCAGATTACAT  
TGGCAGCACTATTCCCGGAAGCGGATGCCCTCAGCAAGCCTGGCAGAGCAGCAACAGCAAAAT  
AACACATCGGGAAAA  
>CGI-74  
GTAGAGAAAGCACGGCAAAGAAAAGAGAAGCAGAGGAAGTTATCGGAATTCTATGCCAGCTTCC  
AGTTTCAGCAGCAAACCTCGAGTCTGTGAAGTCTGCTCGCTATTAGGACTTCATGATAAT  
GACAGACGACTGGCTGATCATTGGGGTAAACTGCACCTGGGATTATTGAAATAAGAGAGAAG  
CTTGAAGAATTAAAGAGAGTCGTAGCTGAGAAGCAGGAGAAAAGAAACCAGGAACGGCTGAAACGA  
AGAGAAGAGAGAGAGAGAGAAGAAAGGGAGAAGCTGAGGAGGTCCCAGTACACAGCAAGAATCCA  
AAAAGG  
>CLH-17  
ATGGCCCAGATTCTGCCAATTGTTTCAGGAGCATCTCCAGCTCCAGAACCTGGGTATCAACCCA  
GCAAACATTGGCTTCAGTACCCGACTATGGAGTCTGACAATTCTGATTAGAGAAAAAGTA  
GGAGAGCAGGCCAGGTGTAATCATTGATATGAATGACCCAAGTAATCCAATTGAAAGACCAATT  
TCAGCAGACAGGCCATCATGAATCCAGCTAGCAAAGTAATTGCACTGAAAGCTGGAAAACCTT  
CAGATTTAACATTGAAATGAAAATGAAGGCTCATACCATGACTGATGATGTACCTT  
TGAAATGGATCTTGAATACGGTTGCTTGTACGGATAATGCAGTTATCACTGGAGTATG  
GAAGGAGAGTCTCAGCCAGTGAAGGAAATGTTGATGCCATTCTAGCCTTGAGGGTGCAGATTATC  
AATTACCGTACAGATGCAAAACAAAAGTGGTTACTTCTGACTGGTATATCTGCACAGCAAAATCGT  
GTGGTGGAGCTATGCAGCTATATTCTGTAGATAGGAAAGTGTCTGAGCCCATTGAAAGGACATGCA  
GCTAGCTTGCACAGTTAAGATGGAAGGAAATGCAAGAAATCAACGTTATTITGTTTGCAGTT  
CGGGGCCAAGCTGGAGGGAAAGTTACATATTATTGAAAGTTGGCACACCCACCTACAGGGAACAGCCC  
TTTCCAAGAAGGCAGTGGATGCTCTTCTCCAGAAGCACAAAATGATTTCTGTTGCAATG  
CAGATCAGTGAAGGACATGATGTTGTTGATAACCAAGTATGGTTATATCCACCTCTATGAT  
CTTGAGACT  
>CLK1  
GACCGCGTGGGTCTGAAACATCTGAATAACACAGACCCAAACAGTACTTCCGCTGTGCCAGATG  
TTGGAATGGTTGAGCATCATGGTCACATTGCAATTGTTGAACATTGGGACTTAGTACTTAC  
GACTTCATTAAAGAAAATGGTTCTACCATTTGCACTGGATCATATCAGAAAGATGGCATATCAG  
ATATGCAAGTCTGTGAATTGGTTCACAGTAATAAGTTGACTCACACAGACTTAAAGCCTGAAAAC  
ATCTTATTGTCAGTCTGACTACACAGAGGCGTATAATCCAAAATAAAACGTGATGAAACGCACC  
TTAATAATCCAGATAATTAAAGTTGAGACTTGGTAGTGCACATATGATGACAAACATCACAGT  
ACATTGGTATCTACAAGACATTATAGAGCACCTGAAGTTATTAGCCTAGGGTGGTCCCAACCA  
TGTGATGTCGGAGCATAGGATGCATTCTTATTGAAATACTATCTGGGTTACCGTATTCCAACA  
CACGATAGTAAGGAGCATTTAGCAATGATGGAAAGGATTCTGGACCTCTACAAAACATATGATA  
CAGAAAACAGGAAACGTAAAATTTTACCCACGATCGATTAGACTGGGATGAACACAGTCTGCC  
GGCAGATATGTTCAAGACGCTGTAACCTCTGAAAGGAATTATGCTTTCTCAAGATGTTGAAACAT  
GAGCGTCTCTTGCACCTCATCAGAAAATGTTGGAGTATGATCCAGCCAAAAGAATTACTCTCAGA  
GAAGCCTTAAAGCATCCTTCTGACCTTCTGAAGAAAAGTATA

Figure 6 (continued)

>DRP-1  
AAGGACAACCTTACCCGTACCGGAGGGTAGAGGAGCGGATGACCGTCGTCGG  
GACAAGGGCGTGGCTACTGGCAAAATGGATGAGAACCGAGTTGTCGCTGTCACCAGCACCAATGCA  
GCCAAGATCTTAACCTGTACCCAAGGAAGGGCGGATTGCCGTGGCTCGGATGCCGACGTGGTC  
ATCTGGGACCCGACAAGTGAAGACCATAACAGCAAAGTCACAAGTCGGCGGTGGAGTACAAC  
ATCTCGAGGTATGGAGTGCCACGGCTCCCCACTAGTGGTCATCAGCCAGGGCAAGATCGTC  
GAAGACGAAACATCAACGTCAACAAGGGCATGGGCCCTCATTCCGCGAAGGCCTTCCCAGG  
CACCTGTACCGCGTCAAAATCAGGAATAAGGTTTGAGTGCAGGGTTCCAGGGCATG  
TATGACGGTCTGTGTACGAGGTACAGCTACACCCAAATATGCAACTCCCCTCAGCCAAA  
TCTTCGCTCTAAACACCAGCCCCACCCATCAGAAACCTCCACCAAGTCAACTTCAGCTTATCA  
GGTGCAGAGATGACAACAATCCCAGGCGACCGGCATCGTGGCGCCCCCTGGTGG  
CGCTCCAACATCACCAGCCTCGGT  
>EF1A  
ATGCACCATGAAGCTTGAGTGAAGCTCTGGGACAATGTGGCTTCATGTCAAGAATGTG  
TCTGTCAAGGATGTCGTGCGAACGTTGCTGGTACAGCAGAAAATGACCCACCAATGGAAGCA  
GCTGGCTTCACTGCTCAGGTGATTATCCTGAACCATCCAGGCCAAATAAGGCCGGCTATGCCCT  
GTATTGGATTGCCACACGGCTCACATTGCATGCAAGTTGCTGAGCTGAAGGAAAAGATTGATCGC  
CGTTCTGGTAAAAGCTGGAAGATGGCCCTAAATTCTGAAGTCTGCTGATGCTGCCATTGTTGAT  
AIGGTTCTGGCAAGGCCATGTGTGAGAGCTCTCAGACTATCACCTTGGGTCGTTGCT  
GTTCGTGTATGAGACAGACAGTTGGGTGTCATCAAAGCAGTGGACAAGAAGGCTGCTGGA  
GCTGGCAAGGTACCAAGTCTGCCAGAAAGCTCAGAAGGCTAAA  
>EF1G(bait)  
GCGGCTGGGACCCGTACACGTATCCTGAAAACGGACGGCCTCAAGGCTCTCATCGCTGCTCAG  
TACAGCGGGGCTCAGGTCCGCGTCTCCGACCCACCTCCATTGGCAAACCAACCGC  
ACCCCTGAATTCTCCGAAATTCTCGCCAGGTCCAGCATTGAGGGTGTGATGGATT  
TGTGTGTTGAGAGCAACGCCATTGCCTACTATGTGAGCAATGAGGAGCTGCCGGAAAGTACTCCA  
GAGGCAGCAGCCCAGGTGGTGCAGTGGGTGAGCTTGCTGATTCGATATAGTCCCCCAGCCAGT  
ACCTGGGTGTTCCCCACCTTGGCATTGCACCAACAAACAGGCCACTGAGAATGCAAAGGAG  
GAAGTGAGGCGAATTCTGGGCTGCTGGATGTTACTTGAAGACGAGGACTTCTGGTGGCGAA  
CGAGTGACATTGGCTGACATCACAGTTGTCACCTGTTGTGGCTCTATAAGCAGGTTCTAGAG  
CCTTCTTCCGCCAGGCCTTCCAATACCAACCGCTGGTCTCTCACCTGCATAACCAGCCCCAG  
TTCCGGGCTGCTTGGCGAAGTGAAACTGTGTGAGAAGATGGCCAGTTGATGCTAAAAAGTT  
GCAGAGACCCAACCTAAAAGACACACCACGGAAAGAGAAGGGTCACTGGGAAGAGAACAGAAG  
CCCCAGGCTGAGCGGAAGGAGGAGAAAAAGGGCTGCCCTGCTCTGAGGAGGAGATGGATGAA  
TGTGAGCAGGCGCTGGCTGCTGAGGCCAAGGACCCCTCGCTCACCTGCCAAGAGTACC  
TTTGTGTTGGATGAAATTAAAGCGCAAGTACTCCATGAGGACACACTCTCTGGCACTGCCATAT  
TTCTGGGACCACTTGATAAGGACGGCTGGTCCCTGGTACTCAGAGTATGCCCTCCCTGAAGAA  
CTCACTCAGACCTTCACTGAGCTGCAATCTCATCACTGGAAATGTCAGCGACTGGACAAGCTGAGG  
AAGAATGCCCTCGCCAGTGTCACTCTTTGGAAACCAACAATAGCAGCTCCATTCTGGAGTCTGG  
GTCTTCCGAGGCCAGGAGCTGCTTCCGCTGAGTCCAGATTGGCAGGTGGACTACGAGTCATAC  
ACATGGCGGAAACTGGATCCTGGCAGCGAGGAGACCCAGACGCTGGTTCGAGAGTACTTTCTGG  
GAGGGGGCCCTCCAGCATGTGGCAAAGCCTTCAATCAGGGCAAGATCTTCAAG

Figure 6 (continued)

&gt;EF1G (prey)

GGGGCTGGGACCCCTGTACACGTATCCTGAAAAGTGGAGGGCCTTCAGGCTCTCGCACCAACCCACTTCCATTGGCAAACCAACCACGC  
 TACAGGGGGCTCAGGTCCGCGTCTCCGACCAACCCACTTCCATTGGCAAACCAACCACGC  
 ACCCCTGAATTCTCCGAAATTCTCCGCGAAGGTCCCAGCATTTGAGGGTGTGATGGATTCTGAGGAGCTGCAGGGAAACTACTCCA  
 TGTTGTTGAGAGCAACGCCATTGCCTACTATGTGAGCAATGAGGAGCTGCAGGGAAACTACTCCA  
 GAGGCAGCAGCCCAGGTGGTGCAGTGGTGAGCTTGCTGATTCCGATATAGTGCAGGGAACTACTCCA  
 ACCTGGGTGTCTCCCCACCTTGGCATCATGCACCAACAAACAGGCCACTGAGAATGCAAAGGAG  
 GAAGTGAGGCGAATTCTGGGCTGCTGGATGCTTACTTGAAGACGAGGACTTTCTGGTGGCGAA  
 CGAGTGACATTGGCTGACATCACAGTTGCTGACCCCTGTTGGCTCTATAAGCAGGTTCTAGAG  
 CCTTCTTCCGCCAGGCCTTCCAATACCAACCGCTGGTCTCACCTGCATTAACCAGCCCCAG  
 TTCCGGGCTGTCTTGGCGAAGTGAAGACTGTGAGAAGATGGCCAGTTGATGCTAAAAGTT  
 GCAGAGACCAACCTAAAAGGACACACCACGGAAAGAGAAGGGTTACGGGAAGAGAAGCAGAAG  
 CCCCAGGCTGAGCGGAAGGAGGAGAAAAGGCGGCTGCCCCCTGCTCTGAGGAGGAGATGGATGAA  
 TGTGAGCAGCGCTGGCTGAGCCAAAGGCAAGGACCCCTCGCTCACCTGCCAAGAGTAC  
 TTTGTGTTGGATGAATTAAAGCGCAAGTACTCCAATGAGGACACACTCTGTGGCACTGCCATAT  
 TTCTGGGAGCACTTGATAAGGACGGCTGGTCCCTGTGGTACTCAGAGTATCGCTTCCCTGAAGAA  
 CTCACTCAGACCTTCATGAGCTGCAATCTCATCAGGAACTGGATGTTCCAGCGACTGGACAAGCTGAGG  
 AAGAATGCCCTGCCAGTGTCACTCTTTGGAACCAACAATAGCAGCTCCATTCTGGAGTCTGG  
 GTCTCCGAGGCCAGGAGCTTGCCTTCCGCTGAGTCCAGATTGGCAGGTGGACTACGAGTCATAC  
 ACATGGCGAAACTGGATCCTGCCAGCGAGGAGACCCAGACGCTGGTCAGAGTACTTTCCCTGG  
 GAGGGGGCCTCCAGCATGTGGCAAAGCCTCAATCAGGGCAAGATCTCAAG

&gt;FEZ1

GGCAACTGCTCTGACACTGAGATCCATGAGAAAGAAGAGGAAGAGTCAATGAGAAGAGTGAAT  
 GATTCCGGTATCAACGAGGAGCCTCTGCTCACAGCAGATCAGGTAAATTGAGGAGATTGAGGAAATG  
 ATGCAGAACCTCCAGACCCCTGAGGAAGAAGAGGAGGTTCTGGAAGAAGAGGATGGAGGAGAAACT  
 TCCTCCCAGGCAACTCGGTCCCTGAGGAGATGCAAGGCAATTGACACAGACCTCAACACAAC  
 TGGCCTATGAACGGCTGAGGCACATGTCGGTCTGAGCTGACCGAGCTGGACCAGGTGGAG  
 GGTGCCATCCGTACTCTCGGAGGAGCTGGTGCAGCAGCTGGCCCGCCGGACGAGCTGGAGTT  
 GAGAAGGAAGTGAAGAACTCCTTATCACGGTGTATTGAGGTTCAAGAACAGCAGAAGGAGCAG  
 CGAGAACTGATGAAAAAGAGGCCAGAGAAAGGGCTGAGGCTGAGGAGCAGCCGGATAGAGAAG  
 GGAAACAGATGCCCTCAAGCGCTTCAGCATGGAAGGCATCTCAACATTCTGAGACTGGCATT  
 CGCCAGACCTTGGCTCCTCAGAAACTGACAAACAGTATCTGAACACAGTCATTCTTACGAGAAG  
 AAAGCCTCTCCCTCAGTGGAAAGACCTGCAAGATGCTGACAAACATTCTTGGCCATGAAGGAG  
 GATAATGAGAAGGTGCCTACTTGCTAACGGACTACATTAAAAGTGTCTGCCCTACC

&gt;G45IP1

ATGGCGTCGAGCGGGGGAGCTAGGGAGTTATTGATCACCACTGCAGAGGGCGGTATGCGAC  
 ACACGGGCCAAATATCGAGAGGGACGACGGCCTCGTGTGAAGGTATATACAATCAATTGGAA  
 TCTCAGTACTTATTAAATACAAGAGTTCTGCTGAGTCAAGGAAATTAGTTGAGCGATT  
 GCTTTATATGGTCAATTGAACAGTACAATGCTCTAGATGAATACCCAGCAGAAGACTTTACTGAA  
 GTTATCTTATTAAATTATGAACTTACAAGTCAAGGACAGCCAAGAGAAAATGGATGAAACAG  
 AGTTTCTCGGTGGATTGCTCATGTTGCTATGCTCCAGAAATTGAAACAGTTGAAGAAACTAGA  
 AAAAAGTACAATGCCAAGGCATATGAGTAAAGACTACTGAAAATAAGACCAATTACGTGACA  
 AAGAAGAAATTGGTTACAGAGCATAAAAGACACAGAGGATTTAGACAAGACTTCCACTCAGAGATG  
 TCTGGATTGGTAAAGCTGCTTGAACACTTCTGCAAGGGAACTCAAATCTTATCTCGTATTCC  
 TGTGAATTGCCATTGTTATTCTCCTCAAAATGATGTTGCTCATCCGGGGACCTGTAGACAGA  
 GCACCAACTCCTCTAAGGATGTTAGAAAACATCATAAACATGGGCATTATAACCACAATGAC  
 TCTTGCGGAAACACAGATAAAACTCTTGAAGGAAACTCTGAGGAGCTGGCCTGCCCTGGTGCACAAAGGCT  
 ATTACGTCTCAGAGGCAGTTGACAGATTATGCCTAGGACAACACAACCTGCAGGAGCGAAAAGA  
 AGAAGAGAAGATGATGTTAAAGCTGGAACATTCTCAAACAAACCCAACTGGTAATGAGATTATG  
 ATTGGACCTCTGTTACAGACATCTCTAAAGTGGATATGCACGATGACTCATTGAATACAACGGCG  
 AATTAAATTGGCATAAAACTTAAAGAGGTAATTCTGTCATCTGTCAGGCCAAAGCCTCCAGAGGACAAGCCA  
 GAAGATGTACATACAAGTCATCATTAAAACAAAGAAGAAGATA

Figure 6 (continued)

&gt;G45IP2

AGGACCTGTATGCCCTATATATTTCTCTGCCTGGAGGCTCTGAAATGTTCCGCATCAGGAAC  
 AATGAGAAAGATGCTGAGTGACAGCCACGGCGTGGAGACCCTCGGGACATCCTGCCAGACACCAGC  
 CTTGGGGGCCATCCTCTCAAAATCATCACGGCCAAGGCTGTCTGAAGCTGCAGGCCGGAAAC  
 GCCGAGGAAGCCGCCCTGTGGAGGGATCTGGTCGCAAAGTCTGGCATCCTACTTGGAGACAGCC  
 GAGGAGGGGGTGACCCCTGGCGGGAGCCTGGATGAAAATGTCAGGAGGTGCTGAAATTGCCACC  
 CGGGAGAATGGCTCCTGCTGAGTACCTGGTGGCTATCCCCATGGAGAAAGGCCTTGACTCCAA  
 GGCTGCTCTGCGCAGGCTGCTCCGGCAGATGGCTTCTCCTTGTACGACCCAAGCTGTGCC  
 TTCTCTGCCCTCTATTACTGTGACATCTGCCACCAAGACGATGCCCTCAGTGAATTCCGGCCAGGATC  
 ATCCACAACGGGACCTCACCAAGCGCCCGATCTGCAGGCAGGCCCTGAAGTTCTGACACAGATC  
 CGGGCCAGCCCTCATCAACCTGCAGATGGTGAACGCGCTCTGTACGAGCATGTGGAGCGGATG  
 CACCTCATTGGGAGGAGACGGGAGCAGCTGAAGCTCCTGGGGATTACCTGGCCTGTGCCGGAGT  
 GGCGCCCTGAAGGAGCTCAGCAAGAGGCTAACACAGGAATTATCTCTTGAATCTCCGCATAGG  
 TTCAGTGTGCTGACCTCAACAGATCGCAGACGGGGTGTATGAAGGATTCTCAAGGCCCTGATT  
 GAATTTCGCTCCCAGCATGCTTACCACTGCGACTGTGCAACCCAGCGCGGCTCATCTGCCAGATC  
 TGCCAGCACCACGACATCATCTCCCTTGAGTTGACACCACAGTCAGGTGTGCCGAGTGCAG  
 ACCGTCTTCCACCAGAGCTGCCAGGCTGTGGTGAAGAAGGGCTGCCCGCTGTGCCGCCGGCGC  
 AAGTACCAAGAACAGAACATTTCGCG

&gt;G45IP3

CCTAACAGGGGCCCTCAGCCCTCTAACATGACCTCCGGCTAGCCATGTGATTCACTCCACTC  
 CATAACGCTCCTCATACTAGGCCTACTAACCAACACACTAACCATATACCAATGATGGCGCGATGT  
 AACACGAGAAAGCACATACCAAGGCCACACACACCCACCTGTCCAAAAAGGCCCTCGATACGGGAT  
 AATCCTATTATTACCTCAGAAGTTTTCTCGCAGGATTTCTGAGCCTTACCACTCCAG  
 CCTAGGCCCTACCCCCAATTAGGAGGGCACTGCCACAGGCTAACCCGCTAAATCCCCT  
 AGAAGTCCCACCTCTAACACATCCGTATTACTCGCATCAGGAGTATCAATCACCTGAGCTACCA

&gt;GADD4.5G

GGTGCAGGGCGCTGAGCCGGATTGGAGTGTGGTGGAGTTGGGAGCCAAGGGTGTGCGCGGTGG  
 CCGGGCTGGGTCTCCGCCGCCCTCCGGCGTCCCGCTCACTGCGCTGGCTCTCCGCAGG  
 ATGCAGGGTGCCTGGAAAGCGCTGCATGAGTTGCTGCTGCGCGAGCGTCAGGGCTGCCACT  
 GCCGGCGTCTACGAGTCAGCCAAAGTCTGAAACGGGACCCGACAATGTGACCTCTGTGTGCTG  
 GCTGCGGGTGAGGAGGGACGAGGGGACATCGCGCTGCAGATCCATTACGCTGATCCAGGGCTTTC  
 TGCTGCGAGAACGACATCGACATAGTGCCTGGCGATGTGCAGGGCTGGGGCTATGTGGGG  
 GCCGGCGAGGAGGGGGTGCGCCGGCGACCTGCAGTCATCCTCATTCGAACCCAACGAGGAC  
 GCCTGGAAGGATCCCCCTTGGAGAACGTCAGCCCTGTTTGCAGGAGAGGCCAGCGTTAACGAC  
 TGGTGCCAGCATCACCCCTCCCCGAG

Figure 6 (continued)

&gt;GIT1

CCACAGATGGCTGACAGATCTGGCAAAAGTGCATGTCAGAGCCTGACTTATCGAATTGGCC  
 AAAGCTGCTAAGAAGAAGCTGCAGGCGCTAGCAACCGCTTTTGAGGAACCTGCCATGGACGTG  
 TATGACGAGGTGGATCGAAGAGAAAATGATGCAGTGTGGCTGGCTACCCAAAACCACAGCACTCTG  
 GTGACAGAGCGCAGTGCCGTGCCCTCCTGCCGTAAACCGGAATACTCAGCCACGCCAATCAG  
 GGGCGACAAAAGCTGGCCCGTTAATGCCGAGAGTTGCCACCTGATCATCGACATTCTCAGT  
 GAGGCCAAGCGGAGACAGCAGGGCAAGAGCCTGAGCAGCCCCACAGACAACCTCGAGCTGTCTG  
 CGGAGCCAGAGTGACCTCGACGACCAACACGACTACGACAGCGTGGCCTTGACGAGGACACAGAC  
 CAGGAGCCCCCTGCGCAGCACCGGCCACTCGGAGCAACCGGGCCGGAGCATGGACTCCTCGGAC  
 TTGCTGACGGGCTGTGACGCTGAGGAGTACCTGGAGCTGAAGAAGGCCCTGGCTACATCGGAG  
 GCAAAGGTGCAGCAGCTCATGAAGGTCAACAGTAGCCTGAGCAGGCTCCGGAGGCTGCAGCGA  
 GAGATCCACAAGCTGCAGGCGGAGAACCTGCAGCTCCGGCAGCCTCCAGGGCCGGTGCACACCT  
 CCACTCCCCAGTGAACGGGCGGAACACACACACCCATGGCGCAGGCGGGAGCACACACCGCAGGGAT  
 CGCCAGGCCTTTCCATGTATGAACCTGGCTTGCCCTGAAGCCCTTGGGGGCCCTGGGAC  
 GAGCTCACTACGGGCTGCAGCTTCCACAGCACTGAGCTAGAGGACGACGCCATCTATTAGTG  
 CACGTCCTGCTGGCCTTACCGGATCCGAAAGGGGTGTCTGCCTCAGCTGTGCCCTTCACTCCC  
 TCCTCCCCGCTGCTGTCTGCCAGGAGGGAAAGCCGCCACACGAGCAAGCTTCCGCCACGGC  
 AGTGGAGCCGACAGTGAATGAGAACACGCCAAAGTGGGGACCCACTGCTGGGCTGGAAGGGAAAG  
 AGTTTCTAGAGCTGGCAAAGAGGAAGACTTCCACCCAGAGCTGAAAGCCTGGATGGAGACCTA  
 GATCCTGGGCTTCCCAGCACAGAGGATGTCATCTTGAAAGACAGAGCAGGTCACCAAGAACATTAG  
 GAACTGTTGGGGCAGCCAGGAGTTCAAGCATGACAGCTTGTGCCCTGCTCAGAGAACATCCAT  
 TTGGCTGTGACCGAGATGGCTCCCTTCCCTAAAGAGGCCAGCCCTGGAGGCCAGTGGAGCTCA  
 CTGGGCTGCTCAACGCCAGGCCCTACCGGCTGCAGAGTGAGTGCCGAAAGACAGTGCCCTAGAG  
 CCCGGGCCCTAGTGGACTTCCAGCTGCTGACTCAGCAGGTGATCCAGTGCCTATGACATGCC  
 AAGGCTGCCAAGCAGCTGGTACCATCACACCCGAGAGAACAGCAG

&gt;hADA3

AAAGATGTGGATGCCCTGCTGAAGAAGTCTGAGGCCAGCATGAACAGCCGAAGATGGATGCC  
 TTTGGTGCCTGACGCAGGCCCTCCTGCAGGCCCTGGTGGAGAAAATATTATTCCTATGGAG  
 GATTCTCCTATTCCCTGACATGTCGGAAAGAATCAGGGCTGACGGGCAAGCACCTCCCCCTGC  
 AATCAGAACAGCCCTCAGTGTGCCCATACTAAGTCCCTGGAGAGGCCCATCAAGGAGGAGCTA  
 ATTGCCAGGGCTTTGGAGTCTGAGGACCGCCCGCAGAGGACTCCAGGATGAGGTCTTGCT  
 GAGCTCGAAACGGCAGGCTGAGCTGAGGCACTTAGTGCCACAACCGCACCAGAACGAC  
 CTGCTGAGGCTGGCAAAGGAGGAGGTGAGCCGGCAGGAGCTGAGGAGCAGGGGTGCGCATGGCTGAC  
 AACGAGGTATGGACGCCCTTCGCAAGATCATGGCTGCCGGCAGAAGAAGCGGACTCCCACCAAG  
 AAAGAAAAGGACCAGGCCTGGAAGACTCTGAAGGAGCGTGAAGAGCATCCTGAAGCTGCTGGATGGG

Figure 6 (continued)

>HB01  
 GACGCTGAGAGGCAGGGCACTAGGGATCGTCCGCAGGATTGGACTGATAACAGAGGCCACG  
 GAGCCCCGGAGCCACCGTCCCTGCTGCCGCTGCCGAATCGAACCGTCGGCCGCAAG  
 CCGCCGGCAATGCCGAGAAGAGGAATGCAGGCAGTAGTTCAAGATGGAACCGAAGATTCCGAT  
 TTTTCTACAGATCTCAGCACACAGACAGTTCAAGAAAGTGTGGCACATCCCAGCATGCTCGA  
 GTCACCCGCTCCTCAGCCAGGCTAAGCCAGAGTTCTAAGATTCCAGTCCTGAAATCTGCAG  
 TCTTTGGCACTGAGGAGGCTGCTTACTCTACCAAGAAGAGTGACCGTAGTCAGCAGCAGCTACC  
 CCAGTGACACCGAAAAAAATACCCCTCTCGGCAGACTCGTTCACTGTTCAAGAAACTGAGCAAGTG  
 GTTGATTTTCAGATAGAGAACTAAAAATACAGCTGATGATGAGTCACCGCCTCGAAGTCCA  
 ACTGGAATGCGCTTCTTCTGAGTCTGACATAGACATCTCAGCCCCAATGTATCTCACGATGAG  
 AGCATTGCCAAGGACATGTCCTGAAGGACTCAGGCAGTGATCTCTCATGCCCAAGCGCCGT  
 CGCTTCCATGAAAGCTACAACCTCAATATGAAGTGTCCTACACCAGCTGTAACTCTCTAGGACAC  
 CTTACAGGAAAACATGAGAGACATTCTCCATCTCAGGATGCCACTGTATCATAACCTCTCAGCT  
 GACGAATGCAAGGTGAGAGACAGAGCCGGATAAGCAGATAGAAGAAAGGATGCTGTACAGG  
 CAAGATGACAACAAAGGCATGCAACCAGGCACCAAGGACACAGGAGAGGAGCTTCGATATAAG  
 GAAAAAGTGGCTGAACTCAGGAAGAAAAGAAATTCTGGACTGAGCAAAGAACAGAAAGAGAAATAT  
 ATGGAACACAGACAGACCTATGGGAACACACGGGAACCTCTTTAGAAAACCTGACAAGCGAGTAT  
 GACTTGGATCTTTCGAAGAGCACAAGCCGGGCTTCAGAGGATTGGAGAACTTAAGGCTGCAA  
 GCCCAAATCACAGAGGGAAAGCAACATGATTAAAACAATTGCTTTGGCCGCTATGAGCTTGATACC  
 TGGTATCATTCTCCATATCTGAAGAATATGCACGGCTGGACGCTCTATATGTTGAAATTCTGT  
 TTAAGGAAATATGAGAGGCCAACGATACTCCGCCGGCACATGGCAAATGTGTGTTGAAACACCCA  
 CCTGGTGTGAGATATATGCAAGGTTCAATCTCTGTGTTGAGTGGATGGCAAGAAAAACAAG  
 ATCTACTGCCAAAACCTGTGCCTGTTGGCAAACCTTTCTGGACGACAAAGACATTATATTATGAT  
 GTGGAGCCCTTCCCTGTTCTATGTTGACAGAGGGGACAACACTGGCTGTCACCTGATTGGATAT  
 TTTTCTAAGGAAAAGAATTCTCCTCAACTACAACGTCCTCTGTATCCTTACTATGCCTCAGTAC  
 ATGAGAÇAGGGCTATGCCAAGATGCTTATTGATTGCTTCAAGTCAAGAAAA  
 GTGGCTCCCCAGAACGTCCACTCTCAGATCTGGGCTTATAAGCTATCGCAGTTACTGGAAAGAA  
 GTACTTCTCCGCTACCTGCATAATTCAAGGAAAGAGATTCTATCAAAGAAATCAGTCAGGAG  
 ACGGCTGTGAATCCTGTGGACATTGTCACTGCAAGGCCCTCAGATGCTCAAATACTGGAAG  
 GGAAAACACCTAGTTAAAGAGAGACAGGACCTGATTGAGTGGATAGCCAAAGAGGCCAAAAGG  
 TCCAACCTCAAATAAACCATGGATCCCAGCTGCTAAATGGACCCCTCCCAAGGGCACT

Figure 6 (continued)

&gt;HD1.7

ATGGCGACCTGGAAAAGCTGATGAAGGCCTCGAGTCCTCAAGTCCTCCAGCAGCAGCAG  
 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG  
 CGCCGCCGCCGCCCTCAGCTCAGCTCAGCCGCCGCCAGGCACAGCCGTGCTGCTCAG  
 CGCAGCCCCCCGCCGCCACCCGGCCGGCTGTGGCTGAGGAGCCGCTGCAC  
 CGACAAAGAAAGAACCTTCAAGCTACCAAGAAAGACC GTGTGAATCAT TGTCTGACAATATGTGAA  
 AACATAGTGGCACAGTCTGTCAAGAAATTCTCAGAAACTCTGGGCATCGCTATGGAA  
 CTTTCTGCTGTGAGTCAGCTCAGATGTCAAGGATGGTGCTGACGAATGCCTCAAC  
 AAAGTTATCAAAGCTTGTATGGATTCTAATCTCCAAGGTACAGCTCGAGCTATAAGGAAATT  
 AAAAGAAATGGTGCCTCGGAGTTGCGTGTGCCCTGTGGAGGTTGCTGAGCTGGCTACCTG  
 GTTCCGGCTCAGAAATGCAGGCCCTACCTGGTAACCTCTGCCGTGACTCGAACAGCAAG  
 AGACCCGAAGAACATCAGTCCAGGAGACCTGGCTGCA GCTGTCCAAAATTATGGCTTCTTGGC  
 AATTTGCAAATGACAATGAAATTAAAGGTTTGTAAAGGCTTCATAGCGAACCTGAAGTCAAGC  
 TCCCCCACCATTGGCGGACAGCGGCTGGATCAGCAGTGA GCACTCGCCAGCACTCAAGAAGGACA  
 CAATATTCTATAGTGGCTACTAAATGTGCTCTTAGGCTACTCGTTCTGTGAGGATGAACAC  
 TCCACTCTGCTGATTCTGGCGTGTGCTCACCCCTGAGGTATTGGTGCCTTGCTGAGCAGCAG  
 GTCAAGGACACAAGCCTGAAAGGCAGCTCGGAGTGACAAGGAAAGAAATGGAAGTCTCCTTCT  
 GCAGAGCAGCTTGTCCAGGTTATGA ACTGACGTTACATCATAACACAGCACCAAGACCACAATGTT  
 GTGACCGGAGCCCTGGAGCTGTTGAGCAGCTCTCAGAACGCCCTCACCCGAGCTTGTGCAAACC  
 CTGACCGCAGTCGGGGCATTGGCAGCTCACCGCTGCTAAGGAGGAGTCTGGTGGCCGAAGCCGT  
 AGTGGGAGTATTGTGAACTTATAGCTGGAGGGGTTCTCATGCA GGCCTGTCCCTCAAGAAAA  
 CAAAAAGGCAAAGTGTCTTAGGAGAAGAAGAACCTTGGAGGATGACTCTGAATCGAGATCGGAT  
 GTCAGCAGCTTGCTTAACAGCCTCAGTGAGGATGAGATCAGTGGAGAGCTGGCTGCTTCTCA  
 GGGTTTCCACTCCAGGGTCAGCAGGTATGACATCATCAGAACAGCCACGGTACAGCACACT  
 GCAGGCGGACTCAGTGGATCTGGCCAGCTG

&gt;Hdd1.0

ATGGCGACCTGGAAAAGCTGATGAAGGCCTCGAGTCCTCAAGTCCTCCAGCAGCAGCAG  
 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG  
 CGCCGCCGCCGCCCTCAGCTCAGCTCAGCCGCCGCCAGGCACAGCCGTGCTGCCCTCAG  
 CGCAGCCGCCCGCCGCCGCCACCCGGCCGGCTGTGGCTGAGGAGCCGCTGCAC  
 CGACCAAAGAAAGAACCTTCAAGCTACCAAGAAAGACC GTGTGAATCAT TGTCTGACAATATGTGAA  
 AACATAGTGGCACAGTCTGTCAAGAAATTCTCAGAAATTCTCAGAAACTCTGGGCATCGCTATGGAA  
 CTTTCTGCTGTGCA GTGATGACGAGTCAGATGTCAAGGATGGTGGCTGACGAATGCCTCAAC  
 AAAGTTATCAAAGCTTGTATGGATTCTAATCTCCAAGGTACAGCTCGAGCTCTATAAGGAAATT  
 AAAAGAAATGGTGCCTCGGAGTTGCGTGTGCCCTGTGGAGGTTGCTGAGCTGGCTCACCTG  
 GTTCCGGCTCAGAAATGCAGGCCCTACCTGGTAACCTCTGCCGTGACTCGAACAGCAAG  
 AGACCCGAAGAACATCAGTCCAGGAGACCTGGCTGCA GCTGTCCAAAATTATGGCTTCTTGGC  
 AATTTGCAAATGACAATGAAATTAAAGGTTTGTAAAGGCTTCATAGCGAACCTGAAGTCAAGC  
 TCCCCCACCATTGGCGCACAGCGGCTGGATCAGCAGTGA GAGCATCGCCAGCACTCAAGAAGGACA  
 CAATATTCTATAGTGGCTACTAAATGTGCTCTTAGGCTTACTCGTTCTGTGAGGATGAACAC  
 TCCACTCTGCTGATTCTGGCGTGTGCTCACCTCTAGA

Figure 6 (continued)

>HDd1.3  
 CCAAGGTTACAGCTCGAGCTCTATAAGGAAATTAAAAAGAATGGTCCCCCTCGGAGTTTCGTGCT  
 GCCCTGTGGAGGTTGCTGAGCTGGCTCACCTGGTCAGAAATGCAGGCCTTACCTGGTG  
 AACCTTCTGCCGTGCCGTGACTCGAACAGCAAGAGACCCGAAGAATCAGTCCAGGAGACCTGGCT  
 GCAGCTGTTCCAAAATTATGGCTTCTTGGCAATTGCAAATGACAATGAAATTAAAGGTTTG  
 TTAAAGGCTTCATAGCGAACCTGAAGTCAAGCTCCCCACCATTGGCGGACAGCGGCTGGATCA  
 GCAGTGAGCATCTGCCAGCACTCAAGAAGGACACAATATTCTATAGTTGGCTACTAAATGTGCTC  
 TTAGGCTTACTCGTCCCTGTCGAGGATGAAACACTCCACTCTGCTGATTCTGGCGTGTGCTCAC  
 CTGAGGTATTGGTGCCTTGCTGAGCAGCAAGGACACAAGCCTGAAAGGCAGCTCGGA  
 GTGACAAGGAAAGAAATGGAAGTCTCTTCTGAGAGCAGCTTGTCCAGGTTATGAACTGAC  
 TTACATCATAACACAGCACCAAGACCAATGTTGTGACCGGAGCCCTGGAGCTGTTGCAGCAGCTC  
 TTCAGAACGCCCTCCACCCGAGCTTCTGCAAACCTGACCGCAGTCGGGGCATTGGCAGCTCAC  
 GCTGCTAAGGAGGAGTCTGGTGGCGAAGCGTAGTGGAGTATTGGAACCTATAGCTGGAGGG  
 GGTTCCCTATGCGAGCCCTGTCCTTCAAGAAAACAAAAGGAAAGTGCCTTACGGAGAAGAAGAA  
 GCCTTGGAGGATGACTCTGAAATCGAGATCGGATGTCAGCAGCTCTGCTTAACAGCCTCAGTGAAG  
 GATGAGATCAGTGGAGAGCTGGCTGCTCTTCAGGGTTTCACTCCAGGGTACGAGGTACATGAC  
 ATCATCACAGAACGCCACGGTACAGCACACTGCAGCAGGACTCAGTGGATCTGGCAGCTG  
 >HDexQ20  
 ATGGCGACCCCTGGAAAAGATGATGAAGGCCTTCGAGTCCCTCAAGTCCTTCCAGCAGCAGCAG  
 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG  
 CCGCCGCCGCCCTCCTCAGCTTCTCAGCCGCCGCCAGGCACAGCCGCTGCTGCCCTCAGCCG  
 CAGCCGCCGCCGCCGCCGCCACCCGGCCGGCTGTGGCTGAGGAGCCGCTGCCACCG  
 CCG  
 >HDexQ51  
 ATGGCGACCCCTGGAAAAGCTGATGAAGGCCTTCGAGTCCCTCAAGTCCTTCCAGCAGCAGCAG  
 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG  
 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG  
 CAACAGCCGCCACCGCCGCCGCCGCCGCCGCTCCTCAGCTTCTCAGCCGCCGCCAGGCA  
 CAGCCGCTGCTGCTCAGCCGAGCCGCCGCCGCCGCCACCCGGCCGGCTGTGGCTGAGGAGCCGCTGCCACCG  
 CTGAGGAGCCGCTGCAACCGACCGTAGTCAG  
 >HIP1  
 GCTGACACCCCTGCAAGGCCACGGGACCGCTTCATGGAGCAGTTACAAAGTGAAGATCTGTC  
 TACCGCTCCAGCAACCTGCAAGTACTTCAGCAGCTCATTCAAGATCCCCAGCTGCTGAGAACCCA  
 CCCAACCTCCTGCGAGCCTCAGCCCTGTAGAACATATCAGCCCTGGGTGATCCCTGAGAG  
 GCTCATCCCCGACAGCGAGCCAGTCCTAGAGAAGGATGACCTCATGGACATGGATGCCCTCAG  
 CAGAATTATTGACAACAAGTTGATGACATCTTGGCAGTTCAATTCAAGCAGTGAATCCCTCAAT  
 TTCAACAGTCAAAATGGTGTGAACAAGGATGAGAAGGACCACTTAATGAGCAGTATAACAGAGAG  
 ATCACTGGATTGAAGGCACAGCTAGAAAACATGAAGACTGAGAGGCCAGCGGGTTGTGCTGAGCTG  
 AAGGGCCACGTCAGCGAGCTGGAGCAGATCTGGCGAGCAGCACCTGGGGAGCAGCGGGGCC  
 GACGACTGTGAATTCTGCGGGCAGAACACTGGACAGCTCAGGAGGAGCAGCGGGAGGACACCGAGAAG  
 GCTCAGCGGAGGCTGCTGAGATAGAAAGGAAAGCTCAAGCCAATGAACAGCGATATAGCAAGCTA  
 AAGGAGAAGTACAGCGAGCTGGTTCAAGAACACACGCTGACCTGCTGCGGAAGAATGCAAGAGGTGACC  
 AACACAGGTGTCATGCCAGACAAGCCAGGTAGATTGGAACGAGAGAAAAAGAGCTGGAGGAT  
 TCGTTGGAGCGCATCAGTGACCAAGGGCCAGCGGAAGACTCAAGAACAGCTGGAAGTTCTAGAGAGC  
 TTGAAGCAGGAACCTGCCACAAGCCAACGGGAGCTTCAGGTTCTGCAAGGAGCAGCTGGAAACTTCT  
 GCCCAGTCAGAACAGCAACTGGGAGCCAGTTGCCAGGCTAGAGAAGGAGCGGGACAGCCTGGTG  
 AGTGGCGAGCTCATAGGGAGGAGGAATTATCTGCTCTCGAAAGAAGACTGCAGGACACTCAGCTC  
 AAACCTGGCCAGCACAGAGGAATCTATGTGCCAGCTGCCAAAGACCAACGAAAAATGCTCTGGTG  
 GGGTCCAGGAAGGCTGCCAGGCTGAGCAGGTGATACAAGACGCGTCAGCGGCCCG

Figure 6 (continued)

&gt;HIP11

GTGGACCTTGTACCGCCTGTGACATCCGGTACTGTGCCAGGATGTTCTTCAGGTGAAGGAG  
 GTGGACGTGGTTGGCTGCCGATGTAGGAACACTGCAGGCCCTGCCAAGGTATCGGGAAACCAG  
 AGCCTGGTCAACGAGCTGGCCTCACCGCCCGCAAGATGATGGCTGACGAGGCCCTGGGAGTGGG  
 CTGGTCAGCCGGTGTCCCAGACAAAGAGGTATGCTGGATGCTGCCCTAGCGCTGGCGGCCGAG  
 ATTCCAGCAAGAGCCCCGTGGCGGTGCAAGACCAAGGTCAACCTGCTGTATTCCCGAACCAT  
 TCGGTGGCCAGAGCCTCAACTACGTGGCGTCTGGAACATGAGCATGCTGCAGACCCAAGACCTC  
 GTGAAGTCGGTCCAGGCCACGACTGAGAACAAAGGAACGTGACCTTCTCCAAGCTC

&gt;HIP13

CCCTGCTGCTCTGAGGACACCATCCCTCCAAGTTTCAAGATTATGATTATTCCTGTAAAGTGGT  
 GACCAAGGAGGACATCAGCAGGAGTTCGACAAGTCCACCAATTCCAAGAAAACAGCGACATCAGC  
 CAGTCCTACCGACGGATGTTCAAGCCAAGCGTCCAGCCTCAACTGCTGGCTCCCCACCAACCTG  
 GGACCTGCTATGGTCACTCCAGGGGTTGCAACTATCCGACGACCCCTCCACCAAGCCTCTGTC  
 CGCCGGGGAACCATGGAGCTGGTCCCATCCCCATCAAGACACCCGTATCCCTGTCAAGACCCCA  
 ACCGTCACAGACCTCCAGGGGTTGCAAGCCCTCCAGATGGGCCAGAACAGCGGGGGAGCAC  
 AGCCCTGAGTCGCATCTGTGGGTGAGGGCCCCAAGGTGTCACCAGCATGCCCTCTCAATGTGG  
 AGCGGCCAGCTCCGTTAACCCCTCCACTTCCAGGCCAGAACGGGATCTCCACAAAGGAGAACATG  
 CAGGCAATTCCAGAAAGTGAAGCTGAAGACCAGGAACGGGACCCCCAAGTGCCACTGTCTCCCCA  
 GGCCAGATTCCAGAGGTGACCCCTGCAGACCTGAGGCCAAGGGATACTCCACAAAGGAGAACATG  
 CTGAACGCCATCCGAAGGGCGTAAACTGAAGAACACGACAAACGATCGCTAGCCCTCGC  
 TTTCT

&gt;HIP15

ATTCACATGGCTCCACCTTATCAAATCTAAACATGATTGAGACATTCAATATGTCAAGTGTGTGAG  
 GAAACCTTGCACATAGTGTGGATTCCCTTGAGCAGCTGACTGGAATAAGGATGCTTAGACACCTC  
 ACTATGACTATTGACTATCACACACTGATTGCCAACTATATGTCCGGTTCTCTCCTTATTAACC  
 ACAGCCAATGCGAGAACGAAGTTCACGTTCTGAAAATGCTATTGAATTGCTCTGAAAATCCTGCT  
 GTGGCAAAAAAAACTATTCACTGCCAAAGCTTTCAATATTGTGGGTCTCTTAAACATAGAAGAG  
 ACAAAATGATAATATTCAAATTGTTATTAAAATGTTCAAGAATATCAGTAACATTATAAAAGTGG  
 AAGATGTCTTAATTGATGATGATTTCAGTCTTGAGGCCCTTATTCTGCATTGCAATTGAG  
 GAGTTAGCTAAGCAACTACAAGCCAAATAGACAACCAAAATGATCCCTGAGGTGGACAACAAAGT

&gt;HIP16

GATGAAGAGGGAGAGAAACATAGGCAGATGATAAAGGAAGCTTTGCTGGGATGATGTACATCAGA  
 GATTCTTGAAGAGAACAGGGAGCTGTGGAGGCCAGTAAGCCAAGGACGTGGACCTGACACTA  
 CCTGGCTGGGCCAGTGGGTGGTGTGGCCTAAAGCCAGTGCCAAGAAAAGACGCCGGTTCTC  
 ATTAAGCCCTGAGGGTCTCCAAGAAAAGATAAGAATTGCAAATGTGATTATCAATGAGAAG  
 CGCAACATCCACGCACTGCTCATCAGGTACGAGTGCTCCATATCCATTACCCACCATGGCAA  
 TTGAAAGGACCATCCAGACCCCCATAGGATCCACATGGAACACCCAGAGGGCTTCCAAAAGCTG  
 ACTACTCCCAAGGTGTCACCAAGCCAGGCCATACTCATTAACCCATAAAAGCAGAACAGTGGGC  
 TACCGGTCTCTCAAGGTGCGACCTGTCATACAGAGGAATCCAAAACGAATCACCACACGT  
 CAACAAAACAGCTGAAGAAATGCTCTGTAGAT

&gt;HIP2

ATGGCCAACATCGCGGTGCAGCGAATCAAGCGGGAGTTCAAGGAGGTGCTGAAGACCGAGGAGACG  
 AGCAAAATCAAATTAAAGTAGATCTGTAGATGAGAATTTCAGAATTAGAGGAGAAATAGCA  
 GGACCTCCAGACACACCATATGAAGGAGGAAGATACCAACTAGAGATAAAAATACAGAAACATAC  
 CCATTAAATCCCCCTAAGGTCCGGTTATCACTAAAATATGGCATCTTAATATTAGTTCCGTACA  
 GGGGCTATTGTTGGATATCTGAAAGATCAATGGCAGCTGCAATGACTCTCCGCACGGTATTA  
 TTGTCATTGCAAGCACTATTGGCAGCTGCAAGGCCAGATGATCCACAGGGATGCTGTAGTAGCAAAT  
 CAGTACAACAAAATCCGAAATGTTCAAACAGACAGCTGACTTTGGGCACATGTGTATGCTGGA  
 GCACCACTTCTAGTCCAGAACACACCAAAATAGAAAACCTATGTGCTATGGCTTGTAGAG  
 AATGCAGTAATAGTGGCTTGTCTCAAAATCATGGATGTAGAGACTGCAACAGAACAGTGTCTG  
 AGTAACGT

Figure 6 (continued)

&gt;HIP5 (bait)

TTTCTTAAAAGTATTTAAAGAAAATCTAAATATGAACATGGTTATCTTAAGGCATTAATTATA  
 AATCAGAGCTTAAGTTGAAATCAAAAGCAGCAGCTATCAGAGATAGTATTGAATTAAACAAAG  
 GAAAAGGTGCAGAAATTCCAAAGACTATTAAAAACTGAGGTGGTTGATGAAACTAGCAATATA  
 GAAAACAATGCTGAAAACAGTCATTCACTGAAGAATAAAACAGGAACAACAGCATTCTCAA  
 CAATTCCACATTCAAAGTGGTGTGAAGCAACATAATTAGTGTCTACTGTGCTGAAATTCT  
 GCTGATACAAAGAAGTCCAGGGAGGATTCTATCTGAAAATGTTACGACTTAGGAGGATCTGGA  
 GCAGACCATATGCCCTTGAACTGTTTATACCTTCAGGTATAACTTGCTAACATGCCCTGGCCA  
 GCCTCAAAAAGAAGAAGTAAAATCCCTGTACATGATGATTCTAAACTAACAGGTAAGCCA  
 CAAAGAGGTAGAGCAAAATAATTAGAAAACCAGGATCTGCAAAGTCAAACATTTACACAAGCT  
 ACAAAACAGAAAAGGCGCTGTCAATTCAACCACAGTCAGTGTGAAAGTCAAACAGGCTTATATGT  
 CAGGGAAAATTAAATTATACCTTGCTCCTCAATCTACATCAAATTAGTAGAAGTGGTAAAAT  
 ATACAAGTGTCTAGTGTCAACCAGTAACCTCTGAAATTCTCAAACATTACACATAACTCT  
 TTTAATTCAAACATGTGCTTCCAACAGAACAGTTGAAATCAGTGGAAATCAGGAAAGTAGTTCT  
 CCACTCTCAAATGCTTGTGAACTGTTCTGACCTAGTCAGTGTGACTCTAAGAAGTCAAAC  
 GAGTGCAGGAAACTTGTGAAATGAGAAAGTCAAACATGCTGAAAGTCAAACAGGCTTATATGT  
 ACATTATATTGCAACCAAGAAGTCTGAGCTGAAAGACAGAAGATCCCTAGAGTCCCTTAATGATCTCAATGAAAGA  
 CTACATTATATAAGAATCCATTGCAAAACCCATCCATCAAACATTACACATAACTTACAAATAATACCA  
 CTTCTGGAGAAGAGAGAAGATAGAACACAGCAGCTGCAGAGACAAGAGA

&gt;HIP5 (prey)

TTTCTTAAAAGTATTTAAAGAAAATCTAAATATGAACATGGTTATCTTAAGGCATTAATTATA  
 AATCAGAGCTTAAGTTGAAATCAAAAGCAGCAGCTATCAGAGATAGTATTGAATTAAACAAAG  
 GAAAAGGTGCAGAAATTCCAAAGACTATTAAAAACTGAGGTGGTTGATGAAACTAGCAATATA  
 GAAAACAATGCTGAAAACAGTCATTCACTGAAGAATAAAACAGGAACAACAGCATTCTCAA  
 CAATTCCACATTCAAAGTGGTGTGAAGCAACATAATTAGTGTCTACTTGCTGAAATTCT  
 GCTGATACAAAGAAGTCCAGGGAGGATTCTATCTGAAAATGTTACGACTTAGGAGGATCTGGA  
 GCAGACCATATGCCCTTGAACTGTTTATACCTTCAGGTATAACTTGCTAACATGCCCTGGCCA  
 GCCTCAAAAAGAAGAAGTAAAATCCCTGTACATGATGATTCTAAACTAACAGGTAAGCCA  
 CAAAGAGGTAGAGCAAAATAATTAGAAAACCAGGATCTGCAAAGTCAAACAGGCTTATATGT  
 ACAAAACAGAAAAGGCGCTGTCAATTCAACCACAGTCAGTGTGAAAGTCAAACATTTACACAAGCT  
 CAGGGAAAATTAAATTATACCTTGCTCCTCTCAATCTACATCAAATTAGAAGTGGTAAAAT  
 ATACAAGTGTCTAGTGTCAACCAGTAACCTCTGAAAATCTCAAACATTACACATAACTCT  
 TTTAATTCAAACATGTGCTTCCAACAGAACACAGTTGAAATCAGTGGAAATCAGGAAAGTAGTTCT  
 CCACTCTCAAATGCTTGTGAACTGTTCTGACCTAGTCAGTGTGACTTCAAGAAGTCAAAC  
 GAGTGCAGGAAACTTGTGAAAATTCAATGGCACTCAAGCAGTGTGCCCCGGCAAGATGCG  
 ACATTATATTGCAACCAAGAAGTCTGTTGAAAGAAGAGGTCTTAATGTCCTGCATCAAATAAGAGGGCT  
 GCTGAAGAAGAATCAGTCCCTTATGAAAAGAGGTCTTAATGTCCTGCATCAAATAAGAGGGCT  
 ACAGGGTCTACTGTATGAGAAGAAAAGCAATTGCTGAAACTAACAGGGAGAAATTTAGAGCAG  
 AAAAGACAAAACCCCTGGATCTGAGGACAGAAGTACAGTGAGCAAATTAAATAATTGGACAAAGT  
 GTCCCTGCTAAGTCAAGTGTGAGGCAAAACAAACTACAAGGGTACTCTTATATTGAAGAAGTTCT  
 GATAGTACTCTGAGTTTGTGATGGCTGAAAACCTAGTGAAGACATCAGTGCAGGAGGATGAGGATT  
 CTGACTGTCTGAAATGCAAACAGATAACAGAAAATCTACCTTAAATAAAACTAACACAAATT  
 AACATCTGCACACTGTCAAGCTGAAGAACAGAACAGATCCCTAGAGTCCCTTAATGATCTCAATGAAAGA  
 CTACATTATATAAGAATCCATTGCAAAACCCATCCATCAAACATTACAAATAATACCA  
 CTTCTGGAGAAGAGAGAAGATAGAACACAGCAGCTGCAGAGACAAGAGA

Figure 6 (continued)

&gt;HMP

CAAGAACAAAGTTAAAATTGAGTCTCTAGCCAAGAGCTTAGAAGATGCTCTGAGGCCAAACTGCAAGT  
 GTCACTCTGCAGGCTATTGCAGCTCAGAAATGCTGCCGTCCAGGCTGTCAATGCACACTCCAACATA  
 TTGAAAGCCGCCATGGACAATTCTGAGATTGCAGGGAGAAGAAATCTGCTCAGTGGCGCACAGTG  
 GAGGGTGCATTGAAGGAACGCAGAAAGGCAGTAGATGAAGCTGCCGATGCCCTCTCAAAGCCAA  
 GAAGAGTTAGAGAAGATGAAAAGTGTGATTGAAAATGCAAAGAAAAAGAGGTTGCTGGGGCCAAG  
 CCTCATATAACTGCTCAGAGGGTAAACTTCACAAACATGATAGTTGATCTGGATAATGTGGTCAA  
 AAGGTCCAAGCAGCTCAGTCTGAGGCTAAGGTTGATCTCAGTATCATGAGCTGGTGGTCAAAGCT  
 CGGGATGACTTAAACGAGAGCTGGACAGTATTACTCCAGAAGTCCTTCTGGTGGAAAGGAATG  
 AGTGTTCAGACTTAGCTGACAAGCTCTACTGATGATCTGAACCTCCCTCATTGCTCATGCACAT  
 CGTCGTATTGATCAGCTGAACAGAGAGCTGGCAGAACAGAAGGCCACCGAAAAGCAGCACATCACG  
 TTAGCCTTGGAGAAACAAAAGCTGGAAGAAAAGCAGGGCATTGACTCTGCAGTAGCAGAAAAGCATT  
 GAACATCACAGAAGTGAATACAGGCTGAACAGGACAGAAAGATAGAAGAAGTCAGAGATGCCATG  
 GAAAATGAAATGAGAACCCAGCTTCGCCACAGGCAGCTGCCACACTGATCATTGCGAGATGTC  
 CTTAGGGTACAAGAACAGGAATTGAGTCTGAATTGAGCAGAACCTGCTCTGAGAAAACCTCTGAA  
 CAAGAATTACAATTTCGTCGTCAGTCAGAGAGCAAGTTGACAACATTACTCTGGATATAAAACT  
 GCCTATGCCAGACTCAGAGGAATCGAACAGGCTGTCAGAGCCATGCAGTTGCTGAAGAGGAAGCC  
 AGAAAAGCCCAACCTCTGGCTTCAGTGGAGGCATTAAAGTACAGCATGAAAGAACCTCATCTGCA  
 GAAACACCTACTATCCCCTGGTAGTGCGGTTGAGGCCATCAAAGCCAACTGTTCTGATAATGAA  
 TTCACCCAAGCTTAACCGCAGCTATCCCTCCAGAGTCCCTGACCCGTGGGGTGTACAGTGAAGAG  
 ACCCTTAGAGGCCGTTCTATGCTGTTCAAAAACAGGCTGTCAGAGCCATGCAATGATTGATGAAACC  
 AGAAAATAGCTTGTACCAAGTACTCCTCTCTACCTACAGTCCTGCTCCTATTCCACCTCAGCAA  
 CTGAAGCCGCCAGAGCTCTGCCCTGAGGATATAAACACATTTAAATTACTGTCATATGCTTCC  
 TATTGCAATTGAGCATGGTGTACGGAGCTAGCAGCAAAGTTGTCATCAGCTGAAGGGGAATCC  
 AGACGAGTGGCACAGGACTGGCTGAAGGAAGCCGAATGACCCCTAGAAACGAAACAGATAGTGGAA  
 ATCCGTACAGCATATGCCAGGCCGTAGGAATAGGAACCACACTCAGTGCAGCCAGAG

&gt;HP28

CCGCCCGCAGACTCTTGCTCAAGTACGACACCCAGTGCTGGTGAGCCGGAACACGGAGAAACGG  
 AGCCCCAAGGCTCGGCTACTGAAAGTCAGCCCCCAGCAGCCCTGGACCTTCAGGTTCAGCCCCACAG  
 CCACCCAAGACCAAGCTCCCCTCAACTCCTGTCAGGCTCAGATCCTACAAAGCAGGCAGAAGAAATC  
 TTGAATGCCATACTACCCCCAAGGGAGTGGGTGGAAGACAGCAGCTATGGATCCAGCAGGTGTCC  
 AGCACCCCTAGCACCAGGATGGACGCTGGTGCACCTCCAGGAGCAGTTAGACTTAAAGCTGCAGCAG  
 CGGCAGGCCAGGGAAACAGGCATCTGCCCTGTCGCCAGGGAACTCTACTCACAGTGTGATGAG  
 TTGATCCGGGAGGTACCATCAACTGTCGGAGAGGGGCTGCTGCTGCCAGTCGGGACGAG  
 ATCCGCATGACCATCGCTGCCCTACCAAGACCCCTGTACAGAGAGCAGCGTGGCGTTGGCATGAGGAAG  
 GCACGTCAGGCTGAGCAGGGAAAGTCAGACATGGAGAGGAAATCGCAGAATTGGAGACGGAAAAG  
 AGAGACCTGGAGAGGAAGTCAGAACGAGCAGAAGGAAAATGTGAAGCCACTGAGAACGGGAGAGC  
 GAGAGGCCAGGTGGAGGGAGAAGAAGCACAATGAGGAGATTCAAGTCCCTGAAGCGAACAAATCAG  
 CAGCTGAAGGCCAACTGGAAAGGCATTATTGCACCAAAAGAAG

Figure 6 (continued)

&gt;HSPC232

CGGCAGGGAGCGGACGGCTGCATTACGGGTCTCCGGAGGGCAGAGTCGGCTTACAGAAGA  
 GACGAAATGTGGTCTGAGGGACGATGAATAATGAAAGAATTCCGAGAGAACGAGCACCTCCTCGA  
 AGTCATCCCAGTGTGAACTGGTTAGATGGACAAGAGACGATCATTCTGCAAGCAGGAAACCT  
 GAATACAGGGACATGAGAGATGGCTTAGAAGAAAAAGTTCTACTCTTCCATTATGCGAGAGAG  
 CGGTCTCCTTATAAAAGGGACAATACTTTTCAAGAGAATCACCTGTTGGCCGAAAGGATTCTCA  
 CACAGCAGATCTGGTCCAGTGTCACTGAGCTACTCTCCAGAAAGGAGCAAATCATACTCT  
 TTCCATCAGTCTCAACATAGAAATAAGAGAGGCTGTCAGTCTTGAAAACATCAAGAGATACT  
 TCACCCCTCAAGTGGTCAGCAGTTCTCATCAAAGGTGTTAGACAACCAGTAGGCTAAGTGA  
 AAGGAACCTGGCTGAGGCTGCAAGCAAGTGGCTGCTGAAAAGCTAGAGAAATCAGATGAAAGTAAC  
 TTGCTGAAATTCTGAGTATGAGGCGGATCCACAGCACCATTGTTACTGACCAGGCCAGAGGAA  
 CCTGAGTCAAACACAACATGGGATAGAATTATTGAAAGATAGTCAGCTAACACTCGCTCTAAA  
 GCAATAGCATCAAAACAAAGAGATTGAACAGGTTACCGACAAGACTGTGAAACTTCCGGATG  
 GTGGTGAAATGCTGATTGAAAAAGATCCTCATTAGAAAAGTCTATACAGTTGCATTGAGGCAG  
 AATTACATGAAATAGGTGAGCGGTGTTGAAGAACTCAAGCATTGCAAGTATGAGTATGATACT  
 TCCACTCAAGATTGGAGAGCCTTT

&gt;HYPA

GGCCGCCGGCGGAGCAGTCTGAGCCCCACGATGAGGCCGGGACGGGAGCTGAGCGTGGAGGCCTC  
 ATGATGGGGCACCCCTGGCATGCATTATGCCCAATGGGAATGCACCCATGGGTAGAGAGCGAAT  
 ATGCCTCTGTACCTCATGGAATGATGCCGCAAGATGATGATGCCCTATGGGAGGGCACCAATGGGA  
 CAAATGCCCTGGAATGATGTCGTCACTAATGCCCTGGAATGATGATGTCATATGTCAGGCTTCC  
 ATGCAGGCTGCCCTACCGCAGGAGTAAATAGTATGGATGTAGCAGCAGGTACAGCATCTGGTCA  
 AAATCAATGTGGACTGAACATAAATCACCTGATGGAAGGACTTACTACTACAACACTGAAACCAA  
 CAGTCTACCTGGGAGAAACAGATGATCTTAAACACCTGCTGAGCAACTCTTATCTAAATGCC  
 TGGAAGGAATAACAAATCAGATTCTGAAAGCCTTACTATTATAATTCTCAAACAAAAGAATCTGC  
 TGGCCAAACCTAAAGAACCTGAGGATCTGAGGATACCAGAATACCATTGCTGGAAGTCTT  
 ATTACAAAATCAAACCTGCATGCAATGATCAAAGCTGAAAGAAGCAGTAAGCAAGAAGAGTCACC  
 ACAACATCAACAGCTCCAGTCCCTACACAGAAATTCCGACCACAATGAGCACCATGGCTGCTGCC  
 GAAGCAGCAGCTGCTGTTGCAGCAGCAGCGGAGCAGCAGCAGCTGCAGCCAATGCT  
 AATGCTTCCACTCTGCTTAAACTGTCAGTGGAACTGTTCCAGTTGCTGAGCCTGAAAGTT  
 ACTTCCATTGTTGCTACTGTTGTTAGATAATGAGAATACTGAGGAAACAAGCA  
 CAACTTACTAGTACCCCTGCTATTCAAGTGTGGAAGTATCCAGTAATACTGGAGAAGAA  
 ACATCTAACAGCAAGAAACTGTAGCTGATTTCCTTAAAGAAGAGGAGAGGCCAACAGCA  
 AAGAAAACATACACTTGGAAATACAAAGGAAGAGGCAAGCAAGCTTTAAAGAATTATTGAAAGAA  
 AACGGGGTACCATCGAATGCTCATGGGAGCAGGCTATGAAAATGATTATAATGATCCACGATAC  
 AGTGCCTTGGCAAAGTTAAGTGAAGAAAAGCAAGCCTTAATGCCTATAAGTCCAGGCAAAAAAA  
 AAAAGAAAAAAAAAAAAAA

&gt;HZFH

CACGCCCGCTTCGCCAGGGCGAGTGCCTGGCGAGAGGCCACCAGCACCTCCAAGGAGTCGCTG  
 GCAGGGAAACAAGCCGCCAACGCCGCTCCCTGCACAAGGTTCTGAAACCAGCTGGAGGAGTTGCTGAGC  
 GACATGAAGGCCGACGTGACCCGCCAGCCAGCACGCTGCCCAGTACCCCCCATCGCAGCCCG  
 CTTCAAGATGTCGAGCGCAGCATCCTCAGCCGCTGGCCAGCAAGGGCACGGAGCCTACCCCCACA  
 CGGGCCTACCCGCCGGTCCCTACGCTACACCTCCGGGTACGGGGCGGCTTCAGCGCCGCACCC  
 GTAGGGGCCCTGGCCGCCAGGCGCCAATTACAGCCAGATGCCTGCAGGGTCTTCACTCACAGCC  
 GCCACCAACGGCCCTCCAGTGTGAAGAAGGAGAAGGAAATGGTGGGGCATTGGTGTAGAC  
 GGGCTGGATCGGAAGGAGCCCCGAGCCGGGAGGTGATCTGTATAAGCAGAC

Figure 6 (continued)

&gt;IKAP

CTCAAAGAAGGCAGTCGCTGGAGGACCTGGCCCTCCTGGAGGCAGTGAGTGAAGTGGTGCAGAAC  
 ACTGAAAACCTGAAAGATGAAGTATAACCATATTTAAAGGTACTCTTCTTTGAGTTGATGAA  
 CAAGGAAGGAAATTACAGAAGGCCCTTGAAGATAACGCTGAGTTGATGAAAGGTCACTCCAGAA  
 ATTTGGACTCTTACTTACCAAGCAGAATTCACTACCCCCGGTTCTAGGTCCAATCTACTGCAAAT  
 AGTATCATGGCATCTTATCAGCAACAGAAAGACTTCGGTTCTGATGCTGAGCTTTTATA  
 CCACCAAAGATCAACAGAAGAACCCAGTGGAAAGCTGAGCCTGCTAGAC

&gt;IMPD2

GACTTCTCATTCTCCCTGGGTACATCGACTTCAGTGAGACCAGGTGGACCTGACTCTGCTCTG  
 ACCAAGAAAATCACTCTTAAGACCCCAGTGGTTCTCTCCATGGACACAGTCACAGAGGCTGGG  
 ATGGCCATAGCAATGGCGCTTACAGGCGGTATTGGCTTATCCACCACAACGTACACCTGAATT  
 CAGGCCAATGAAGTTCGGAAAGTGAAGAAATATGAACAGGGATTCACTCACAGACCCGTGGTCTC  
 AGCCCCAAGGATCGCGTGCGGGATGTTTTGAGGCCAAGGCCGGATGGTTCTGCGGTATCCA  
 ATCACAGACACAGGCCGGATGGGAGCCGCTTGGTGGCATCATCTCCCTCAGGGACATTGATT  
 CTCAAAGAGGAGGAACATGACTGTTCTGGAAGAGATAATGACAAAGAGGAAAGACTTGGTGGTA  
 GCCCCTGCAGGCATCACACTGAAGGAGGAAATGAAATTCTGCAGCGCAGCAAGAAGGGAAAGTTG  
 CCCATTGTAATGAAGATGATGAGCTTGTGGCATCATTGCCGGACAGACCTGAAGAAGAATCGG  
 GACTACCCACTAGCCTCAAAGATGCCAAGAACAGCTGCTGTGGGGCAGCCATTGGCACTCAT  
 GAGGGATGACAAGTATAAGGCTGGACTTGCTGCCAGGCTGGTGGATGTTGACTCT  
 TCCCAGGGAAATTCCATCTTCCAGATCAATATGATCAAGTACATCAAAGACAAATACCTAATCTC  
 CAAGTCATTGGAGGCAATGTGGTCACTGCTGCCAGGCCAGGCAAGAACCTCATTGATGCAGGTGTGGAT  
 GCCCTGCGGGTGGGCATGGGAAGTGGCTCATCTGCATTACGCAGGAAGTGCCTGGCCTGTGGGGCG  
 CCCCAAGCAACAGCAAGTGTACAAGGTGTAGAGTGCAGGCCAGGCTGGTGTCCGGTATTGCT  
 GATGGAGGAATCCAAAATGTGGGTATTTGCGAAGGCCCTGGCCCTACAGTCATG  
 ATGGGCTCTCTCCTGGCTGCCACCAACTGAGGCCCTGGTGAATACTCTTCCGATGGGATCCGG  
 CTTAAAGAAAATATCGCGGTATGGTTCTCGATGCCATGGACAAGCACCTCAGCAGGCCAGAACAGA  
 TATTTCACTGAGCTGACAAAATCAAAGTGCAGGCCAGGAGTGTCTGGCTGTGCAGGACAAAGGG  
 TCAATCCACAAATTGTCCCTACCTGATTGCTGCCATCCAACACTCATGCCAGGACATTGGTGC  
 AAGAGCTGACCCAAAGTCCGAGCCATGATGACTCTGGGAGCTTAAGTTGAGAAGAGAACGTCC  
 TCAGCCCAGGTGGAAGGTGGCGTCCATAGCCTCATTGATGAGAAGCGGCTTTTC

&gt;KPNA2

GCTTGGGCACTCACTAACATTGCTTCTGGGACATCAGAACAAACCAAGGCTGGTAGATGGAGGT  
 GCCATCCCAGCATTCTCATTCTGTGTTGGCATCTCCCCATGTCACATCAGTGAAACAGCTGTCTGG  
 GCTCTAGGAAACATTGCAAGGTGATGGCTCAGTGTCCGAGACTTGGTTATTAAAGTACGGTGCAGTT  
 GACCCACTGTTGGCTCTCCTGCAAGTCTCTGATATGTCATCTTAGCATGTGGTACTTACGTAAT  
 CTTACCTGGACACTTTCTAATCTTGCCTGCCAACAGAACATCTGCACCCCGATAGATGCTGTTGAG  
 CAGATTCTCCTACCTTAGTTGGCTCTGCATCATGATGCCAGAACAGTGTAGCAGATACCTGC  
 TGGGCTATTCTACCTTACTGATGGTCAAATGACGAAATTGGCATGGTGGAAAACAGGAGTT  
 GTGCCCAACTTGTGAAGCTCTAGGAGCTCTGAATTGCCAATTGTGACTCTGCCCTAAGAGCC  
 ATAGGGAATATTGTCACTGGTACAGATGAAACAGAACACTCAGGTGTGATTGATGCAGGAGCACTCGCC  
 GTCTTCCAGCCTGCTACCAACCCAAAACATCACATTGAGAACAGCTACGTGGACAATGTCA  
 AACATCACAGCCGGCGCCAGGACAGATAAGCAAGTGTGAATCATGGATTAGTCCCATTCCCT  
 GTCAGTGTCTCTCTAAGGAGATTAAAGACACAAAGGAAGCTGTGGGCCGTGACCAACTAT  
 ACCAGTGGTGGAACAGTGTGAACAGAACAGATTGTGACCTTGTACTGTGGCATAATAGAACCGTTGATG  
 AACCTCTTAACGCAAAAGATACCAAGATTATTCTGGTTATCCTGGATGCCATTCAAATATCTT  
 CAGGCTGCTGAGAAACTAGGTGAAACTGAGAAACTTAGTATAATGATGAGAAGATGTGGAGGCTTA  
 GACAAAATTGAGCTCTACAAACCATGAAAATGAGTGTGATGAGGCTCTGTTAAGCTTAATT  
 GAGAAGTATTCTGAGAGGAAGAGATCAAAACGTTGATCAGGAAACTACCTCTGAAAGGC  
 TACACTTCCAAGTTCAAGGATGGGCTCTGGACCTTAACCTT

Figure 6 (continued)

&gt;KPNB1

TGGCAGCTGTGGCTTAGTGGAGACTGTGCCGTGCCCTGCAATCCAACATCATACTTTCTGT  
 GACGAGGTGATGCAGCTGCTCTGGAAAATTGGGAATGAGAACGTCACAGGTCTGTGAAGCCG  
 CAGATTCTGTCACTGTTGGTATATTGCCCTGCTATTGGAGGAGGTTAAAAAATACTTAGAG  
 GTTGTATTGAATACTCTCAGCAGGCCTCCCAGCCAGGTGGACAAGTCAGACTATGACATGGTG  
 GATTATCTGAATGAGCTAAGGGAAAGCTGTTGGAAGCCTATACTGGAATCGTCCAGGGATTAAG  
 GGGGATCAGGAGAACGTACACCCGGATGTGATGCTGGTACAACCCAGAGTAGAATTATTCTGTCT  
 TTCATTGACCACATTGCTGGAGATGAGGATCACACAGATGGAGTAGTAGCTTGTGCTGGACTA  
 ATAGGGACTTATGTACAGCTTGGAGGATGTACTGAAATTAGTAGAAGCTAGGCCAATGATC  
 CATGAATTGTTAATGAGGGGGAGATGAAAGACTAACAAAGCAAACCTTGCTACATGGCA  
 ACAAAAGAACTGAGGAAACTGAAGAACCAAGCT

&gt;Ku70

AAGACCCGGACCTTAATACAAGTACAGGCGGTTGCTTCTGCCCTAGCGATACCAAGAGGGTCTCAG  
 ATCTATGGGAGTCGTCACTTAACTGGAGAACAGAGGAAACAGAACAGCTAAAACGGTTTGATQAT  
 CCAGGTTGATGCTCATGGGTTCAAGCCGGTGGTACTGCTGAAGAAACACCATTACCTGAGGCC  
 TCCCTGTCGTGACCCAGAGGAGTCGCTGGTATTGGGAGCTCACCCCTGTCAGTGCCTGCTC  
 ATCAAGTGTCTGGAGAAGGAGGTTGCAGCATTGTCAGATACACACCCCGCAGGAACATCCCTCCT  
 TATTTGGCTTGGCACAGGAAGAGTGGATGACCAGAAAATTCAAGGTGACTCCTCCA  
 GGCTTCCAGCTGGCTTTTACCCCTTGCTGATGATAAAAGGAAGATGCCCTTACTGAAAAAAATC  
 ATGGCAACTCCAGAGCAGGTGGCAAGATGAAGGCTATCGITGAGAAGCTCGCTCACATACAGA  
 AGTGACAGCTTGAGAACCCGTGCTGCAGCAGCAGTCAAGAACCTGGAGGCCTGGGCTTGGAT  
 TTGATGGAGCCGAAACAAGCAGTGGACCTGACATTGCCAAGGTTGAAGCAATGAATAAAAGACTG  
 GGCTCTGGTGGATGAGTTAACGGAGCTGTTACCCACCAGATTACAATCCTGAAGGGAAAGTT  
 ACCAAGAGAAAACAGATAATGAAGGTTCTGGAAGCAAAAGGCCAAGGTGGAGTATTAGAAGAG  
 GAGCTGAAGACCCACATCAGCAAGGGTACGCTGGCAAGTTCACTGTGCCATGCTGAAGAGGCC  
 TGCCGGCTTACGGCTGAAGAGTGGCTGAAGAAGCAGGAGCTGCTGGAGGCCACCAAGCAC  
 TTCCAGGAC

&gt;LUC7B1

GTCGACGCCGTGCCGTGACGCCGCCGGTTCTGCAAAGGCAGAAAAGTACATGAGTTAAAT  
 GAAAAAAATAGGAAAACCTCCTGCTAAAGCCGAAACAGCTAGGGCTGAAGGTAATGTGGATGAATCC  
 CAGAAGATTCTTATGGAAGTGGAAAAGTTCTGCGGAAGAAAAAGAAGCTGAGGAAGAATACAGA  
 AATTCCATGCCTGCATCCAGTTTCAAGCAGCAGGAAAGCTGCGTGTGCGAGGTCTGCTAC  
 CTTGGTCTCCATGACAATGACCGTCGCGCTGGCAGACCACTTCGGTGGCAAGTTACACTGGGTT  
 ATTCAAGATCCGAGAGAACGCTGATCAGTTGAGGAAAAGTGTGCTGAAAGCAGGAGAACAGAAAAT  
 CAGGATCGCTTGAGGAGGGAGAGGGAGGGGAGCGCTGAGCAGGAGGTGGGATCA  
 AGAACCGAGAGATCCGAGGAGGTACGCTCCGGGATCGCGTGGAGGCGGTCAGATCTACCTCC  
 CGAGAGCGACGGAAATTGTCGGGCTCCGGTCCCGAGATAGACATCGCGCCACCGCAGCGTTCC  
 CGGAGCCACAGCCCCGGACATCGTGGGCTTCCGGGACCGAAGTGCAGAAATACAAGTCTCCAGA  
 GAGCGGGCATCCAGAGAGGGAGTCTGGAGAGCGGGGGCGGAGCGAGCGAGGGCCCCCGGACTGGAGG  
 CTTGAGAGCTCCAACGGGAAGATGGCTTACGGAGGTCAAGAGAGAACGGAGGCCGGAGATC

&gt;MAGEH1

GCATCCTCCCTAGGACTGCTGTAAGCTTGTAGCAGGAGACATGCCCTGGGAGCAAAG  
 AGTCGGCCGCCGTAATGCGAGGCCGAGAACGCAACAATCGAAAATCCAGGCC  
 GAGGCCCTCGAGACCCCTATGGCCGCTCTGTGGTAGCGAGCACCCCGAAGACGACCTGAGCGGC  
 CCCGAGGAAGACCCGAGCACTCCAGAGGAGGCCTTACCAACCCCTGAAGAAGCCTCGAGCACTGCC  
 CAAGCACAAGCCTTCAGTCCCCGGAGCAATTTCAGGGCACCAAGAAAAGTCTCCTGATGTCT  
 ATATTAGCGCTCATCTCATGGCAACAGCGCAAGGAAGCTCTGGTCTGAAAGTGCTGGGG  
 AAGTTAGGAATGCGAGCCTGGACGTCAAGCAGCATCTTGGAGATCCGAAGAACGATCGTACAGAA  
 GAGTTTGCGCAGAGGGTACCTGATTATAAACCGGTGCCCGTAGCAGTCCGGTGGAGTATGAG  
 TTCTCTGGGGCCCCGGAGCACGTTGAATCGAGCAAACGTGAAAGTCATGCAATTGTGGCAAGG  
 GTTGTGTAACCGATGCTCTAAAGACTGGCCTTGTAAATTAGACTGGGATTGGACGATGAGCAGAG  
 GTTGTGAGGCTATCCTCAATTAGGTGCTAGGGTTATTCCGCCCCCT

Figure 6 (continued)

>MAP11c3  
CAGCGGGAGCTCGCCGACCGCTGTAAAGGAGGTACAGCAGATCCGGACCAGCACCCCAGCAAA  
ATCCCGGTGATCATCGAGCGTACAAGGGTGAGAAGCAGCTGCCGTCTGGACAAGACCAAGT  
TTGGTCCCGGACCATGTCAACATGAGCGAGTTGGTCAGAGATCATCCGGCGCCGCTGCAGCTGAAC  
CCCACCGAGGCCTCTCCTGCTGGTGAACCAGCACAGCATGGTGAGTGTGTCCACGCCATCGCG  
GACATCTACGAGCAGGAGAAAGACGAGGACGGCTTCTCTATATGGTCTACGCCCTCCCAGGAAACC  
TTCGGCTTC  
>mHAP1  
CCGAAAGAGCAGGTGCAGAGCGGTGCCGGAGACGGGACAGGGTCGGGGACCCAGCAGCAGGCACC  
CCCACGACCCAGCCTGCAGTTGGTCCCGCTCCGGAGCCCTCGCCGGAGCCAAACCTGCTCAGCG  
CAGGGAACCGGGTCCGGACAAAAATCAGGATCCCAGCAAGACAGGAAGCTTTGTGCGGTCCATG  
ATCATTGGTATTGGACGCACCATGGACCCGCTACGTATTCCAGGGCCTTACGGTCCCCGGCC  
ACTGGCCTGGGCACTGAAAGGCCAGGGAACTGGAAGACACCAGCCGTAACATCGGCCGGAGG  
CCCGCGTGTCCGGCTTGAGCGTGCAGCGTTATTGAGAGCTGCAGGAAGCGTTGTCTTAAT  
CCACCAACCCACGAAGAAGATCACCGAAGATGATGTCAAAGTGTGTTGCTGGAAAGAGAAA  
GAACGGGACCTGAACACAGCCGGATCGGCCAGTCCCTGGTGAACACAGAACAGTGTCTGATG  
GAGGAGAATAATAAGCTGAAACCATGCTGGCTCAGCCAGGGAGGAGATTTACATCTCCGGAAG  
CAGGTGAACCTGCAGGAGATGACCTTCTCAGCTCTACTCAGACTCTGATGACGATGATGAGGAA  
GACGAGGAAGACGAGGAAGAGGGCGAAGAGGGAGGAACGAGAAGGACAGAGGATCAAGACCCAGCAG  
CACGACCACCCCTATGGTCCCCAAGCCACACCTAAGGCTGAGACAGCCACCGCTGCCACAG  
CTGAAACCTGAGCAGGCTCAGGCTCTGGAGGAAGAGAACGACACCTGGAGAGGAGGCC  
TCCACCTGACAACCTGGAGGACGAAGAGCAGATGCTCATTCTGGAATGTTGGAGCAGTTCTCT  
GAAGCCAGCCAGCAGATGGCAGAGCTATCGGAAGTGTGTTGAGGCTGGAAGGCTATGAGAGG  
CAGCAGAAAGAGATCACTCAGCTGCAGGCCAGATCACCAGCTACAAACAGCGTGTGTCAGTCTTAT  
GGGGCCCGAGCGAGGAAACTGCAGCAGATGCTGGCTCAGAGAACGGGATCCACTCGGAGAGCCTG  
CGAGCTGGCTCCTACATGCAGGATTATGGGAGCAGGCCCTCGTGAACGCCAGGAGGATGGGAAAGAGT  
CATGCCAGCGCTCCTCCATGCCCGCAGGCTCTGTCACCCACTATGGATAAGTGTGCTCTGGAT  
GCACTTCAAGTTCCAGAGACACTGGCTGAGGAGCTCCGAACATCTGTGAGGAAGTTCATCACT  
GACCTGCGTATTCATGGAGGAGACGTGACACTCACTGCAGGGAGGGCGTAAGAAGGAGCAGAGG  
GCGATGCCACCCCCACCGCTCA  
>mp53  
GTCACCGAGACCCCTGGCCAGTGGCCCCCTGCCACTCCATGGCCCCCTGTCATCTTGT  
CCTCTCAAAAAAACTTACCAAGGCAACTATGGCTTCCACCTGGCTCTGCACTGGACAGCC  
AAGTCTGTTATGTGCACGTACTCTCTCCCTCAATAAGCTATTCTGCCAGCTGGCGAAGACGTGC  
CCTGTGCAGTTGTGGTCAGGCCACACCTCCAGCTGGAGCCGTGTCGCCATGGCCATCTAC  
AAGAAGTCACAGCACATGACGGAGGCTGAGACGCTGCCACCATGAGCGCTGCTCCGATGGT  
GATGGCTGGCTCCTCCCCAGCATCTTATCGGGTGGAAAGGAAATTGTATCCGAGTATCTGGAA  
GACAGGCAGACTTTGCCACAGCGTGGTACCTTATGAGCCACCCGAGGCCGCTCTGAGTAT  
ACCACCATCACTACAAGTACATGTGTAATAGCTOCTGCATGGGGGATGAACCGCCGACCTATC  
CTTACCATCATCACACTGGAAGAGACTCCAGTGGAACCTTCTGGACGGGACAGCTTGAGGTTCGT  
GTTGTGCTGCCCTGGGAGAGACCGCCGTACAGAAGAAATTCCGAAAAAGGAAGTCCTT  
TGCCCTGAACTGCCAGGGAGCGCAAAGAGAGCGCTGCCACCTGCACAAGGCCCTCTCCCCG  
CAAAAGAAAAAAACCACTTGATGGAGAGTATTCACCCCTAAGATCCGCGGGCGTAAACGCTTCGAG  
ATGTTCOGGGAGCTGAATGAGGCTTAGAGTTAAAGGATGCCATGCTACAGAGGAGTCTGGAGAC  
AGCAGGGCTCACTCCAGCTACCTGAAGACCAAGAAGGCCAGTCTACTTCCGCCATAAAAAAAACA  
ATGGTCAAGAAAGTGGGGCCTGACTCAGAC

Figure 6 (continued)

&gt;NAG4

CGAGACCGGGTGGAGAATGAGGCAGAAAAAGATCTCCAGTGTACGCCCTGTGAGATTAGACTTG  
CCTCCTGAGAACGCTCTCACAGCTTTAGCAAACAAGAAGAAGTAGAACAGACACCCCTCAA  
GAAGCTTGAATCAACTGATGAGACAATTGAGAGAAAAGATCCAAGTGCTTCTTCATTCCT  
GTGACTGATTTATTGCTCTGGCTACTCCATGATCATTAAACACCCAATGGATTTAGTACCATG  
AAAGAAAAGATCAAGAACAAATGACTATCAGTCATAGAACAACTAAAGGATAACTTCAAACAAATG  
TGTACTAATGCCATGATTTACAATAAACAGAGACCATTATTATAAGCTGCAAAGAACGCTGTTG  
CACTCAGGAATGAAAATTCTAGCCAGGAAAGAACATTGAGCCTGAAGCAGAGCATAGACTTCATG  
GCTGACTTGCAGAAAACTCGAAAGCAGAAAGATGGAACAGACACCTCACAGAGTGGGGAGGACGGA  
GGCTGCTGCCAGAGAGAGAGAGGACTCTGGAGATGCCGAAGCACACGCCCTCAAGAGTCCCAGC  
AAAGAAAATAAAAGAACAAAGATATGCTGAAAGATAAGTTAAAAGCAATAATTAGAGAGA  
GAGCAGGAGCAGCTGACCGCATCGTAAGGAATCTGGAGGAAAGCTGACCAGGCGGCTGTGAAC  
AGTCAGTGCAGATTGAAAGAACCCAGATGGAACACGACGTTGGGACTCTCCATCCTGTG  
GATCCCATTGTAGGAGAGCCAGGCTACTGCCCTGTGAGACTGGGAATGACAACGGAAAGACTTCAG  
TCTGGAGTGAATACTTGCAGGGGTTCAAAGAGGATAAAAGGAACAAAGTCACTCCAGTGTATAT  
TTGAATTATGGGCCCTACAGTTCTATGCACCGCATTATGACTCCACATTGCAAATATCAGCAAG  
GATGATTCTGATTTAATCTATTCAACCTATGGGAAGACTCTGATCTCCAAAGTGATTTCAGCATIC  
CATGAGTTTGGCCACGTGCCAAGATTATCCGTATGTCATGGCAGATAGTTACTGGATGTTTA  
ACAAAAGGAGGGCATTCCAGGACCTACAAGAGATGGAGATGTCATTGCCCTGAAGATGAAGGCCAT  
ACTAGGACACTTGACACAGCAAAAGAACATGGAGATTACAGAACAGTAGAGGCCACCAGGGCTTGGAC  
TCCAGTACTCAAGACAGGCTCATAGCGCTGAAAGCAGTAACAAATTGGCGTTCCAGTTGAAGTT  
TTGACTCTGAAAGAGCTGAAATATTCCAGAAAGAAACTTGTGAGACCCAGATTGCTCAGGGAA  
CTCCAGGAAGCCAGAATGAAACGTTGAGCACCAGACCCCCCTCCGAACATGATCTGTCCTTGGGT  
CCCTCATACAGAGAAATGCATCTGCTGAACAAGTGACCAATAATCTTAAAGAACCTTGACAGCAA  
GTAACCTCAGGTGATATCGTAAGCACGTATGGAGTTGAAAGAACACTGAAAGAACCTAAAAAGACGGATGTT  
CCCGTCATGGAAAACAACCTTGTGGATTGACAGAACACTGAAAGAACCTAAAAAGACGGATGTT  
GCTGAGTGTGGACCTGGTGGAAAGT

Figure 6 (continued)

>NEFL

CTCTCTCCCTGCTCTCTCCACCGCCGCCGGGGAGCCACCGGCCGCC  
 ACCATGAGTTCTCAGCTACGAGCGTACTACTCGACCTCTACAAGCGGCCTACGTGGAGACG  
 CCCGGGTGCACATCTCCAGCGTGCAGCGCTACAGCACCGCACGCTCAGCTTACTCCAGCTAC  
 TCGGCGCCGGTGTCTCCTCGCTGTCCGTGCCGCCAGCTACTCCAGCTCTGGATCGTTGATG  
 CCCAGCTGGAGAACCTCGACCTGAGCCAGGTAGCCGCATCAGCAACGACCTCAAGTCCATCCGC  
 ACGCAGGAGAACGGCGAGCTCCAGGACCTCAATGACCGCTCGCCAGCTTCATCGAGCGCGTGCAC  
 GAGCTGGAGCAGCAGAACAAAGTCCCTGGAAGCCGAGCTGCTGGTGTGCCAGAAGCACTCCGAG  
 CCATCCCCTCCGGCGCTGTACGAGCAGGAGATCCGCCACCTGCCCTGGCGGGAAAGATGCC  
 ACCAACGAGAACAGCGCTCCAGGGCGAGCGCGAAGGGCTGGAGGAGACCTGCGCAACCTGCAG  
 GCGCGCTATGAAGAGGAGGTGCTGAGCCCGAGGACGCCAGGGCCGGCTGATGGAAGCGCGCAA  
 GGCGCCACGAGGCCGCGCTCGCTCGCCCGAGCTCGAGAACGCCATCGACAGCTTGATGGACGA  
 ATCTCTTCTGAAGAACAGAACAGAGGAGATCGCCGAACCTGAGGCCAGATCCAGTACGCG  
 CAGATCTCCGTGGAGATGGACGTGACCAAGCCGACCTTCCGCCGCTCAAGGACATCCGCCG  
 CAGTACGAGAACAGCTGGCCGCAAGAACATGCGAGAACGCTGAGGAATGGTTCAAGAGCCCTTCACC  
 GTGCTGACCGAGAGCCCCGCAAGAACACCGACGCCGTGCGGCCCAAGGACGAGGTGTCGAG  
 AGCCGTCGTCTGCTCAAGGCAAGACCCCTGGAATCGAACGATGCCGGGGCATGAATGAAGCGCTG  
 GAGAAGCAGCTGAGGAGCAAGCAGAACGCCACATCGCGCTATGCAGGACACGATC  
 AACAAATTAGAAAATGAATTGAGGACCAAAAGAGTGAATGGCAGCATACTAAAAGAACATACAA  
 GACCTCTCAACGTGAAGATGGCTTGATATTGAGATTGAGCTACAGGAAACTCTTGGAGGC  
 GAGGAGACCCGACTCAGTTCAACCAGCTGGGAAGCATAACCACTGGTACTCTCCAGAGCTCCAG  
 GTCTTGGCGATCTGCCCTACGGCGTTTACAGACCGACTCCCTATCTGATGTCACCCGCTCCTC  
 CGTCCCTACTACACCAGCCATGTCCAAGAGGAGCAGATCGAAGTGGAGGAAACCATTGAGGCTG  
 AAGGCTGAGGAACCCAAGGATGAGCCCCCTCTGAAGGAGAACGCCAGGGAGGAGAACAG  
 GAAGAGGCCGAGGAAGAGGGAGGAGCTGAAGAGAACAGGAGGAGAACAGGAGAACAG  
 AAAGAAGAACAGAACAGGAGGTGAAGGTGAAGAACAGGAGAACACAAAGAACAGTGAAGAGGAGGAG  
 AAGAAAGTTGAAGGTGCTGGGAGGAACAAGCAGCTAACAGAACAG  
 >p53

ATGGAGGAGCCGAGTCAGATCCTAGCGTAGGCCCCCTGTAGTCAGGAAACATTTAGACCTA  
 TGAAACTACTTCTGAAAACAACGTTCTGCCCCCTGCCGCCAGCAATGGATGATTGATG  
 CTGCCCCGGACGATATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCAGAATG  
 CCAGAGGTGCTCCCCCTGTCATCTCTGCCCCCTGCCAGCAGCTCCTACACCGGGCCCTGCCACCGCC  
 CCTCTGGCCCCCTGTCATCTCTGCCCCCTGCCAGAAAACCTACCAAGGGCAGCTACGGTTCCGT  
 CTGGGCTCTTGCAATTGGACAGCCAAGTCTGTGACTTGCACGTACTCCCTGCCCTAACAAAG  
 ATGTTTGCCAACCTGGCCAAGACCTGCCCTGTGAGCTGTGGTTGATTCCACACCCCGCCCGC  
 ACCCGCGTCCGCGCCATGGCCATCTAACAGCAGTCACAGCACATGACGGAGGTGTGAGGCGCTGC  
 CCCCACCATGAGCGCTGCTCAGATAGCGATGGCTGGCCCCCTCCAGCATCTTATCCGAGTGGAA  
 GGAAATTGCGTGTGGAGTATTGGATGACAGAACACTTTGACATAGTGTGGTGGTGCCTAT  
 GAGCCGCTGAGGTGGCTCTGACTGTACCAACCATCCACTACAACATGTGTAACAGTCCCTGC  
 ATGGGGCGCATGAACCGGAGGCCATCTCACCATCATCACACTGGAAAGACTCCAGTGGTAATCTA  
 CTGGGACGGAACAGCTTGAGGTGGCTGTTGTGCCCTGTCCTGGGAGAGACCGGGCGCACAGAGGAA  
 GAGAATCTCCGCAAGAACAGGGAGCCTACCCAGGAGCTGCCCTGGGAGAGACCGGGCGCACAGAGGAA  
 CCCAACACACCAGCTCTCCCCAGCAAAGAACAGAACACTGGATGGAGAATATTCACCCCTT  
 CAGATCCGTGGCGTGAGCGCTCGAGATGTTCCGAGAGCTGAATGAGGCCCTGGAACCTAAGGAT  
 GCCCAGGCTGGGAAGGAGCCAGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGT  
 CAGTCTACCTCCGCCATAAAACTCATGTTCAAGACAGAACAGAAGGGCTGACTCAGAC

Figure 6 (continued)

>PFN2  
 GCTCCTGCCGTCCCGCTGCAGTGCAGAGGGCTCGAAGATGGCCGTTGGCAGAGCTACGTGGAT  
 AACCTGATGTGCCATGGCTGCTGCCAGGAGGCCATTGTCGGCTACTGCAGGCCAAATACGTC  
 TGGGCAGCCACGGCCGGGGCGTCTTCAGAGCATTACGCCAATAGAAATAGATATGATTGTAGGA  
 AAAGACCGGGAAAGGTTCTTACCAACGGTTGACTCTGGCGGAAGAAATGCTCAGTGATCAGA  
 GATAGTCTATACTCGATGGTACTGCACAATGGACATCCGGACAAAGACTCAAGGTGGGAGCCA  
 ACATACAATGTGGCTGCGAGAGCTGGTAGAGTCTGGTCTTGTAAATGGAAAAGAAGGGTC  
 CATGGAGGGCGGATTGAATAAGAAGGCATACTCAATGGCAAAATACTTGAGAGACTCTGGGTT  
 >PIASy (bait)  
 CTGGTGGAGGCCAAAAACATGGTGTGAGTTTCGAGTCTCCGACCTTCAGATGCTCTGGTTTC  
 GTGGGCCGGAGTAAGAGTGGACTGAAGCACGGACTCGTACCCAGGGCCCTCAGCTGGTGCAGTT  
 GACTGTAGCCCTGAGCTGTCAGAAAGATCAAGGAGCTGTACGAGACCCGCTACGCCAAGAAC  
 TCGGAGCCTGCCCCACAGCCGACCGGCCCCCTGGACCCCTGACCATGCACCTCACGACCGG  
 GCCGGCGCTGTGCCAGGACTCCGCTGGCAGGCCAATATTGACTACCCGTGCTACGGAAAG  
 TACTTAAACGGACTGGGACGGTTGCCGCCAAGACCCCTCAAGCCAGAAGTCCGCTGGTGAAGCTG  
 CCGTTCTTAATATGCTGGATGAGCTGCTGAAGCCCACCGAATTAGTCCCACAGAACACGAGAAAG  
 CTTCAAGGAGAGCCCGTGCATCTCGCATTGACGCCAAGACAGGTGGAGTTGATCCGGAACCTCAGG  
 GAAC TGCAAGCCCGAGTTAAAGCCGTGCAGGTGTCCTGAGAATCTGTTACTCAGACACCAAGCTGC  
 CCTCAGGAGGAGCCAGTACCCGCCAACATCGCTGTGAAGGTCAACCACAGCTACTGCTCCGTC  
 GGCTACTACCCCTCCAATAAGCCGGGTGGAGGCCAAGAGGCCGTGCCGCCCATCAACCTCACT  
 CACCTCATGTACCTGTCTCGGCCACCAACCGCATCACTGTCACCTGGGGAACTACGGCAAGAGC  
 TACTCGGTGGCCCTGTACCTGGTGCAGCTGACCTCATCGGAGCTGCTGCAGAGGCTGAAGACC  
 ATTGGGTTAAAGCACCCGGAGCTGTGCAAGGCACTGGTCAAGGAGAAGCTGCGCCTTGATCCTGAC  
 AGCGAGATGCCACCACCGGTGTGCGGGTGTCCCTCATCTGTCGCTGGTGAAGATGCGGCTCTCC  
 GTGCCCTGCCGGCAGAGACCTGCGGCCACCTGCACTGCTTCGACGCCGTCTTACCTGAGATG  
 AACGAGAAAGGCCACCTGGATGTGCCCGTGTGCAGAAGCCAGGCCACCGACTCATC  
 ATCGACGGGCTCCTCTCGAAGATCCTGAGCGAGTGTGAGGACGCCAGGAGATCGAGTACCTGGTG  
 GACGGCTCGTGGTGCCTCGATCCCGCGCGAAAAGGAGCGCAGCTGCAGGCCGCCAGGGCGCCATCCTC  
 GTGCTGGGCCCTCGGACGCCAATGGCTCCTGCCGCCAGCGTCAACGGGAGCGGTGCCCTG  
 GGCAGCACGGGTGGCGGGCCCGGTGGCAGGATGGAGAATGGGAAGGCCGGCGCCATGTGGTG  
 GACCTCACGCTGGACAGCTCATCGTCTCGGAGGATGAGGAGGAGGAGGAAGAGGAGGAGGAAGAC  
 GAGGACGAAGAGGGCCCCCGGCCAACCGCCGCTGCCCTTCCAGAAGGGCTGGTGCGGCCCTGC

Figure 6 (continued)

&gt;PIASy (prey)

CTGGTGGAGGCCAAAACATGGTATGAGTTTCGAGTCTCGACCTTCAGATGCTCTGGTTTC  
 GTGGGCCGGAGTAAGAGTGGACTGAAGCAGCAGCTCGCACCAAGGGCCCTCAGCTGGTGAGTT  
 GACTGTAGCCCTGAGCTTTCAAGAAGATCAAGGAGCTGTACGAGACCCCTACGCCAAGAAC  
 TCGGAGCCTGCCACAGCCGACCCGGCCCTGGACCCCTGACCATGCACCTCCACCTACGACCG  
 GCCGGCGCTGTGCCAGGACTCCGCTGGCAGGCCAAATATTGACTACCCCGTGTCTACGGAAAG  
 TACTTAAACGGACTGGGACGGTTGCCGCCAAGACCCCTCAAGCCAGAAGTCCGCTGGTGAAGCTG  
 CCGTTCTTAATATGCTGGATGAGCTGTGAAGCCCACCGAATTAGTCCCACAGAACAAACGAGAAAG  
 CTTCAAGGAGAGCCCGTGCATCTTCGATTGACGCCAAGACAGGGAGTTGATCCGGAACTCCAGG  
 GAACTGCAGCCCCGAGTTAAAGCCGTGCAGGTGCTGAGAACTGTACTCAGACACCAGCTGC  
 CCTCAGGAGGACCAGTACCCGCCAACATCGTGTGAAGGTCAACCACAGCTACTGCTCCGTCCCC  
 GGCTACTACCCCTCCAATAAGCCGGGGTGGAGGCCAAGAGGCCGTGCCGCCATCAACCTCACT  
 CACCTCATGTACCTGTCCCTGCCACCAACCGCATCTGTCACCTGGGGAACTACGGCAAGAGC  
 TACTCGGTGGCCCTGTACCTCGTGCAGCTGACCTCATCGGAGCTGCTCAGAGGCTGAAGACC  
 ATTGGGGTAAAGCACCCGGAGCTGTCAAGGCACTGGTCAAGGAGAAGCTGCGCCTTGATCCTGAC  
 AGCGAGATGCCACCACCGGTGTGCGGGTGTCCCTCATCTGTCCGCTGGTGAAGATGCCGCTCTCC  
 GTGCCCTGCCGGCAGAGACCTGCQCCCACCTGCAGTGCTTGACGCCGTCTTCTACCTGCAGATG  
 AACGAGAAGAACCCACCTGGATGTGCCCCGTGTGCGACAAGCCAGCCCCCTACGACCCAGCTCATC  
 ATCGACGGGCTCTCTCGAAGATCCTGAGCAGTGAGGACGCCAGGAGATCGAGTACCTGGTG  
 GACGGCTGTGGTGCCTGATCCGCCGAAAAGGAGCGCAGCTGCAGCAGGCCAGGGGCCATCTC  
 GTGCTGGGCCCTCGGACGCCAATGGCTCTGCCGCCAGCGTCAACGGGAGCGGTGCCCTG  
 GGCAGCACGGGTGGCGGCCGGCCGGTGGCAGCATGGAGAATGGGAAGCCGGGCCGATGTGGTG  
 GACCTCACGCTGGACAGCTCATCGTCTCGGAGGATGAGGAGGAGGAAGAGGAGGAGGAAGAC  
 GAGGACGAAGAGGGCCCCGCCAACGCCGCTGCCCTTCCAGAAGGGCTGGTGCCGCCCTGC

&gt;PLIP

GGGGAGATAATCGAGGGCTGCCCTACCGTGCTGCCGGAACCGAGAACAGAACGAGATGAGTGG  
 CCCCTGGCCGAGATCCTGAGCGTGAAGGACATCAGTGGCCGGAAAGCTTTCTACGTCCATTACATT  
 GACTTCAACAAACGTCTGGATGAATGGGTGACGCATGAGCCGCTGGACCTAAAGAAGATCCAGTC  
 CCCAAGAAAGAGGCCAACGACCCCTAAGAACGGACTCCTGGGTCCCGTCTGGCTCTCAGAG  
 AGAGAGCTGAAACGGAAGGTGGAGGTGGTTTACCAAGCAACTCCAGTGGCCAGGCCAGGACGGAAAGCGA  
 GCCTCGGTTTCCCAGAATGGACCCGCCGTAGGGCAGTGGCAGGCCAGGACGGACGGAAAGCGA  
 AAATCGAATTGTTGGCACTGATGAGGACTCCAGGACAGCTCTGATGGAATACCGTCAGCACCA  
 CGCATGACTGGCAGCCTGGTGTCTGATCGAAGCCACGACATCGTACCCGGATGAAGAACATT  
 GAGTCATTGAGCTGGGCCGACGCCCTCAAGCCGTGGTACTTCTCCCGTACCCACAGGAACCTC  
 ACCACATTGCCGTCTCTACCTGTGCGAGTTCTGCCCTCAAGTACGGCCGTAGTCTCAAGTGTCTT  
 CAGCGTCATTGACCAAGTGTGACCTACGACATCCTCCAGGAATGAGATTACCGCAAGGGCACC  
 ATCTCCCTCTTGAGATTGACCGTAAGAACAAAGAGTTATTCCAGAACCTGTGTCTTGGCC  
 AAGTGTCTCTTGACCATAAAGACACTGTACTATGACACAGACCCCTTCCCTCTACGTCTACGACA  
 GAGTATGACTGTAAGGGCTCCACATCGTGGGCTACTCTCCAAAGGAGAAGAACGAGAC  
 TACAATGTGGCCTGCATCTAACCTGCCCTACCGGCCGGCTACGGCAAGCTGCTGATC  
 GAGTTCAGCTATGAACTCTCAAAGTGGAGGGAAAACAGGGACCCCTGAGAACGCCCTCAGAC  
 CTTGGCCTCTATCTATCGAAGCTACTGGTCCAGACCATCCTGGAGATCCTGATGGGCTGAAG  
 TCGGAGAGCGGGGAGAGGCCACAGATCACCATAATGAGATTAGTGAATCACCAGCATCAAGAAC  
 GAGGATGTCATCTCACTCTGCAGTACCTCAATCTCATCAACTACTACAAGGGCCAGTACATCCTC  
 TCCAAGTGTCTGCACTCCAAAGGACTGGAGCAAGAGGGGGAGTG

Figure 6 (continued)

&gt;PTN

TTGAGTCAAAGGCAGGATCAGGTTCCCGCCTTCAGTCAAAAATCCCGCCAAGAGAGCCCCAGA  
 GCAGAGGAAAATCAAAGTGGAGAGAGGGGAAGAAAGAGACCGAGTCAGTCATCCGTCCAGAAGGCG  
 GGGAGAGCAGCAGCGGCCAAGCAGGAGCTGCAGCGAGCCGGTACCTGGACTCAGCGGTAGCAAC  
 CTCGCCCTTGCAACAAAGGCAGACTGAGCGCCAGAGAGGACGTTCCAACTCAAAATGCAGGCT  
 CAACAGTACCAGCAGCAGCGTCGAAAATTGAGCTGCCCTCTTGCATTCAACTGGCA  
 GCTGTGGATACTGCTGAAGCAGGGAAAGAAAGAGAAACAGAAAAAAAGTGAAGAAGTCTGACTGT  
 GGAGAATGGCAGTGGAGTGTGTGCCCACCACTGGAGACTGTGGCTGGGACACCGGAGGGC  
 ACTCGGACTGGAGCTGAGTGAAGCAAACCATGAAGACCCAGAGATGTAAGATCCCCTGCAACTGG  
 AAGAAGCAATTGGCGCGGAGTGCRAAATACCAAGTCCAGGCCCTGGGAGAATGTGACCTGAACACA  
 GCCCTGAAGACCAAGAACCTGGAAAGTCTGAAGCGAGCCCTGACAATGCCGAATGCCAGAAGACTGTC  
 ACCATCTCCAAGCCCTGTGGCAAACGTACCAAGCCAAACCTCAAGCAGAATCTAAGAAGAAGAAA  
 AAGGAAGGCAAGAACAGGAGAAGATGCTGGAT

&gt;PTPK

AGTAACATACATCAATGCTGCTTATGGACAGCTACAGGCAACCAAGCTGCTTCATCGTCACACAA  
 TACCCCTGCCAAACACTGTAAAAGACTTCTGGAGATTAGTGTATGATTATGGCTGTACCTCCATT  
 GTGATGTTAACGAAGTCGACTTGTCCCAGGGCTGCCCTCAGTACTGCCAGAGGAAGGGATGCTA  
 CGATATGGCCCCATCCAAGTGGATGTATGTTCAATGGACTGTGATGTGATCAACCGGATT  
 TTTAGGATATGCAATCTAACAGACACAGGAAGGTTATCTGATGGTCAACAGTTCACTGAGTACCTA  
 GGATGGGCTTCTCATCGAGAAGTGCCTGGATCCAAAAGGTCAATTCTGAAACTGATACTTCAGGTG  
 GAAAAGTGGCAGGAGGAATGCGAGGAAGGGGAAGGCCGACGATTATCCACTGCCAAATGGTGGC  
 GGGCGAAGTGGCATGTTCTGCTATAGGCATCGTGTGAAATGGTGAACAGGCAAACATGGTGGAAAGCCCGGAGCAA  
 GATGTTTCCATGCACTAAAGACACTGAGGAACAGCAAGCCAAACATGGTGGAAAGCCCGGAGCAA  
 TACCGTTCTGCTATGATGTAGCTTGGAGTACCTGGAATCATCT

&gt;SETBD1

AAGGCCCTCACCTCAGGACTAGGCATCAAGGATGAGGGAGACATCAAACAGGCAAGAAAGAGGAC  
 ACTGACGACCGAAACAAGATGTCAGTAGTTACTGAAAGCTCTGAAATTACGGTTACAATCCTCT  
 CCTGTGAAGCCTGAAAGGACTTCGCCGCCACCTAGTAAGACTAGTATGCATCAAAGCCGAAAGACTC  
 ATGGCTTCTGCTCAGTCAACCCCTGATGATGTCTGACACTGTCCAGCAGCACAGAAAGTGAGGG  
 GAAAGTGGGACCAGCCGAAAGCCACTGCTGGTCAGACTCGGCTACAGCGGTTGACAGTGTGAT  
 ATCCAGACCATATCCTCTGGCTCTGAAAGGGATGACTTGTGAGGACAAGAAGAACATGACTGGTCCA  
 ATGAAGCGTCAAGTGGCAGTAAAATCAACCCGAGGCTTGTCTTAAATCAACCCATGGGATTGCA  
 ATTAAATCAACCAACATGCCCTCTGTTGACAAGGGGGAGAGCGCACCTGTTGTAAGAACACACCC  
 CAATTCTATGATGGCGAGGAGTCTTGTACATATTGATGCCAAGCTTGAAGGCAACCTGGCCCG  
 TACCTCAACCACAGTTGCAGCCCCAACCTGTTGTCCAGAAATGTCTTGTGGATACCCATGATCTT  
 CGCTTCCCTGGTGGCCTTCTTGCAGCAAAGAATCCGGCTGGACAGAAACTTACTTGGGAC  
 TACAACCTACGAGGTGGGCAGTGTGGAAGGAGCTACTCTGTTGCTGGGGCATTGAATGCA  
 AGAGGACGTCTTCTT

Figure 6 (continued)

&gt;SH3GL3

GTGGCCGGGCTGAAGAACGAGTCCACAAAGCCAGCCAGCTATTAGTAAAAATAAGTGGTGCT  
 GAAGGAACTAAACTAGACATGAATTCTTGACATGGAAAGGAAAATAGATGTTACCAATAAAGTT  
 GTTCAGAAATTCTTCAAAAACCCTGAATATCTTCAGCCAAATCCAGCATACAGAGCTAACGTA  
 GGAATGCTGAACACTGTGCGAGATCCGAGGGCAGGTGAAGACCAACAGGATACCCGAGACGGAA  
 GGCTTGCTGGGGACTGTATGCTGAAATACGGGAAGGAGCTCGGGGAAGGACTCCACCTTGGCAAT  
 GCATTGATAGAAGTGGTGAATCCATGAAGCTAATGGTGAAGGAGACTCTCTTGATATTAAT  
 GTAAAGCAAACTTTATTGATCCACTTCAGTACTACAAGATAAAAGATTAAAAGAGATCGGGCAT  
 CACCTGAAAAGCTGGAAGGCCGCCCTGGATTACGATTATAAAAAGAACGAGTAGGTAAGATA  
 CCAGACGAAGAAGTCAGACAAGCGGTAGAAAATTGAAGAGTCAGGAGTTGGCTGAAAGAAC  
 ATGTTAACCTTTAGAAAATGATGAGAACAGTCAGCCAGTGGCTGTGTTAGAGGAGCA  
 TTAGACTATCACAGACAGTCCACAGAGATTCTGCAGGAGCTGCAGAGCAAGCTACAGATGCGAATA  
 TCAGCTGCATCCAGTGTCCCCAGACGAGAATACAAGCCAAGGCCGTGAAAAGGAGTTAGTGG  
 CTCAATGGAGTTCCACCACCTCTGAGTGAAGACGACAGGCTAACATTCCATGGACCAGCCC  
 TGCTGCGTGGTCTATGACTTGAGCCAGAAAACCAAGGAGAATTAGGATTAAAAGAAGGGAC  
 ATCATACATTAACCAATCAAATAGATGAAAATGGTATGAAGGAATGATACACGGAGAATCGGGA  
 TTCTCCCCATTAATTACGTGGAAGTGTGCGCTTACCTCAG

&gt;SUMO-2

CGGCCCCGCGCACAGTTCGGCGGGAGAGCGCCGGGGCGAGAGCGTGACTCGCCCGCTCCGCGCT  
 GCTTCCCCCGCGCCGCTCCCCCGCGCTCGCGCAGCCATGTCGAGGAGAACGCCAAGGAGGGT  
 GTGAAGACAGAGAAATGACACATCACCTGAAGGTGGCCGGCAGGACCGCTCCGTGGTGCAGTTC  
 AAGATCAAGAGGCACACGCCGCTGAGCAAGCTGATGAAGGCCTACTGCAGAGGAGGCAGGGCTTGTCA  
 ATGAGGCAGATCAGATTCAAGGTTGACGGCAGCCAATCAATGAAACTGACACTCCAGCACAGCTG  
 GAGATGGAGGACGAGGACACCATCGACGTGTTCCAGCAGCAGACGGGAGGTGTGCCGGAGAGCAGC  
 CTGGCAGGGCACAGTTTC

&gt;SUMO-3

CCCTCGTCCACCGCTGCCGCTCCTCTTCTGCGCCTCTGGTGTGCTTGTGCTCGTTGGTG  
 CGGACCTGGTACCTCTTGTGAAGCGGCAGCTGAGGAGACTCCGGCGCTGCCATGCCAGCAGAA  
 AAGGCCAAGGAAGGAGTCAGACTGAGAACACGATCATATTAAATTGAAGGTGGGGGGAGGAT  
 GGTCTGTGGTGCAGTTAAGATTAAGAGGCATACACCACTTAGAAACTAATGAAAGCCTATTGT  
 GAACGACAGGGATTGTCAATGAGGCAGATCAGATTCCGATTGACGGCAACCAATCAATGAAA  
 GACACACCTGCACAGTTGGAATGGAGGATGAAGATAAAATTGATGTTCCAACAGCAGACGGGA  
 GGTGTCTAC

&gt;TAL1

AGCTCACCCGTGAAGCGTCAGAGGA TGGAGTCGCGCTGGACCAGCTAACAGCAGTTCACCAACCGTG  
 GTGGCCGACACGGCGACTTCCACGCCATCGACGAGTACAAGCCCCAGGATGCTACCAACCAACCG  
 TCCCTGATCCTGCCGAGCACAGATGCCCTTACCAAGGAGGACAGATTAAAAATGCTATTGATAAA  
 ACTTTGTGTTG  
 GGCAGGAAAGCTGGCGGGTACAAGAGGACCAAGATTAAAGGAGCTACCCACTTAGAAACTAATGAA  
 AGCCTATTG  
 GGTCTGTGGTGCAGTTAAGATTAAGAGGCATACACCACTTAGAAACTAATGAAAGCCTATTGT  
 GAACGACAGGGATTGTCAATGAGGCAGATCAGATTCCGATTGACGGCAACCAATCAATGAAA  
 GACACACCTGCACAGTTGGAATGGAGGATGAAGATAAAATTGATGTTCCAACAGCAGACGGGA  
 GGTGTCTAC  
 AGCAAGGACCGAATTCTTAAAGCTGTCACTAACCTGGAAAGGAATTCAAGGCTGGAAAGGAGCTC  
 GAGGAGCAGCACGGCATCCACTGCAACATGACGTTACTCTCTCCGCCAGGCTGTGGCTGT  
 GCCGAGGCGGGTGTGACCCATCTCCCATTTGTTGGCGCATCTTGATGGCATGTGGCAA  
 ACCGACAAGAAATCCTATGAGCCCCCTGGAAGAACCCCTGGGTAAAGAGTGTCACTAAATCTACAA  
 ACTAAGAAGATTAGCTACAAAACCATTGTCATGGCGCCTCTCCGAAACACGGCGAGATC  
 AAAGCACTGGCGGGCTGTGACTTCCCTCACCATCTCACCAAGCTCTGGGAGAGCTGCTGCAGGAC  
 AACGCAAGCTGGTGCCTGTGCTCTCAGCCAAGGCCGGCCAAGCCAGTGACCTGGAAAAAA  
 ATCCAC  
 CTGGATGAGAAGTCTTCCGTTGGTGCACAAACGAGGACCAGATGGCTGTGGAGAAGCTCTGAC  
 GGGATCCGAAGTTGCCGCTGATGCAAGTGAAGCTGGAGCGGATGCTGACAGAACGAATGTTCAAT  
 GCAGAGAATGGAAAG

Figure 6 (continued)

>TCPG  
 CAGACTGACATTGAGATTACCGAGAGGAGGACTTCACCGAATTCTCCAGATGGAGGAAGAGTAC  
 ATCCAGCAGCTGTGAGGACATTATCCAACGTAAAGCCCCATGTGGTCATCACTGAAAAGGGCATC  
 TCAGATTAGCTCAGCACTACCTTATGCCGGCCAATATCACAGCCATCCGCAGAGTCCCGAAGACA  
 GACAATAATCGCATTGCTAGGCCGTGGGGCCCGGATAGTCAGCCGACCAGAGGAAGTGGAGAGAA  
 GATGATGTTGGAACAGGGAGCAGGCCGTGGAAATCAAGAAAATTGGAGATGAATACTTACTTTC  
 ATCACTGACTGCAAAGACCCCAAGGCCGTGACCCATTCTCCTCCGGGGCTAGCAAAGAGATTCTC  
 TCGGAAGTAGAACGCAACCTCCAGGATGCCATGCAAGTGTGCGCAATGTTCTCCTGGACCCCTCAG  
 CTGGTGCCAGGGGGTGGGGCCTCGAGATGGCTGTGGCCCAGTGCCTTGACAGAAAATCCAAGGCC  
 ATGACTGGTGTGGAACAATGCCATACAGGGCTGTGCCCAGGCCCTAGAGGTCAATTCTCGTAC  
 CTGATCCAGAACTGTGGGCCAGCACCATCCGTCTACTTACCTCCCTCGGGCCAAGCACACCCAG  
 GAGAACTGTGAGACCTGGGTGTAATGGTGAAGACGGGTACTTGGTGGACATGAAGGAACGGGC  
 ATATGGGAGCCATTGGCTGTGAAGCTGCAGACTTATAAGACAGCAGTGGAGACGGCAGTCTGCTA  
 CTGCGAATTGATGACATCGTTAGGCCACAAAAAGAAAGGCCATGACCAGAGCCGGCAAGGCCGG  
 GCTCCTGATGCTGGCCAGGAG  
>VIM  
 TCCCCCGGCCAGAGACGCGAGCCGCTCCCACCCACACCCACCGCGCCCTCGTTCGCCCTTTC  
 TCCGGGAGCCAGTCGCCGCACCGCCGCCAGGCCATGCCACCCCTCCGCAGCCATGTCCACC  
 AGGTCCGTGTCCTCGTCCCTACCGCAGGATGTTGGGGCCCGGACCCCGAGCCGGCCGAGC  
 TCCAGCCGGAGCTACGTGACTACGTCACCCGCACCTACAGCCTGGCAGGCCGCTGCCGCCCAGC  
 ACCAGCCGAGCCCTACGCCCTCGTCCCCGGCGGTATGCCACCCGCTCCTCTGCCGTGCGC  
 CTGGGAGCAGCGTGGGGCTCCTGCAGGACTCGGTGGACTTCCTCGTGGAGCTGCAAGGAGCTGAATGACCGC  
 ATCAACACCGAGTTCAAGAACACCCGACCAACGAGAAGGTGGAGCTGCAAGGAGCTGAATGACCGC  
 TTGCGCAAACATACGACAAGGTGCGCTTCTGGAGCAGCAGAATAAGATCCTGCTGCCGAGCTC  
 GAGCAGCTCAAGGGCCAAGGCAAGTCGCCCTGGGGACCTCTACGAGGAGGAGATGCCGGAGCTG  
 CGCCGGCAGGTGGACCAGCTAACGACAAGCCCGCTCGAGGTGGAGCGCAGAACCTGGCC  
 GAGGACATCATGCCCTCCGGAGAACATTGCAAGGAGATGCTCAGGAGGAGATGCTTCAGAGAGAGGAAGCCAAA  
 ACCCTGCAATCTTCAGACAGGATGTTGACAATGCGTCTCTGGCACGTCTGACCTTGAAACGAAA  
 GTGGAATCTTGCAAGAAGAGATTGCCCTTTGAAAGAAACTCCACGAAGAGGAATCCAGGAGCTG  
 CAGGCTCAGATTAGGAACAGCATGCCAAATCGATGTTGATGTTCCAAGCCTGACCTCACGGCT  
 GCCCTGCGTGCAGCTACGTCAAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAAGGAGGAGAAGAA  
 TGGTACAATCCAAGTTGCTGACCTCTGAGGCTGCCAACCGGAACAAATGACGCCCTGCCAG  
 GCAAAGCAGGAGTCCACTGAGTACGGAGACAGGTGCAAGTCCCTCACCTGTGAAGTGGATGCCCTT  
 AAAGGAACCAATGAGTCCCTGGAACGCCAGATGCCGTGAAATGGAAGAGAACCTTGCCGTTGAAGCT  
 GCTAACTACCAAGACACTATTGCCCTGCAAGGAGATGAGATTCAGAAATATGAAGGAGGAATGGCT  
 CGTCACCTCGTGAATACCAAGACCTGCTCAATGTTAAGATGCCCTTGACATTGAGATTGCCACC  
 TACAGGAAGCTGCTGGAAGGCCAGGAGAGCAGGATTCTCTGCCCTTCCAAACTTTCCCTCCCTG  
 AACCTGAGGAAACTAATCTGGATTCACTCCCTCTGGTTGATACCCACTAAAAAGGACACTTCTG  
 ATTAAGACGGTTGAAACTAGAGATGGACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTT  
 GAA

Figure 6 (continued)

>VIMc  
 CAGGAGGGAGATGCTTCAGAGAGAGGAAGGCCAAAACACCCCTGCAATCTTCAGACAGGATGTTGAC  
 AATCGTCTCTGGCACGTCTGACCTTGAAACGCAAAGTGGAAATCTTGCAAGAAGAGATTGCCTT  
 TTGAAGAAAATCCACGAAGAGGAAATCCAGGAGCTGCAGGCTCAGATTCAAGAACAGCATGTCAA  
 ATCGATGTGGATGTTCCAAGCCTGACCTCACGGCTGCCCTGCGTGAACGTACGTCAAGCAATATGAA  
 AGTGTGGCTGCCAAGAACCTGCAGGAGGCAGAAGAAATGGTACAATCCAAGTTGCTGACCTCT  
 GAGGCTGCCAACCGGAACAATGACGCCCTGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGA  
 CAGGTGCACTCCCTCACCTGTGAAGTGGATGCCCTTAAAGGAACAATGAGTCCCTGGAACGCCAG  
 ATGCGTGAATGGAAGAGAACCTTGCCGTTGAAGCTGCTAATACCAAGACACTATTGGCCGCTG  
 CAGGATGAGATTCAAATAAGGAGGAATGGCTGTCACCTCGTGAATACCAAGACACTGCTC  
 AATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGGAGGAGAG  
 AGGATTCTCTGCCCTTCCAAACTTTCCCTGAAACCTGAGGGAAACTAATCTGGATTCACTC  
 CCTCTGGTTGATAACCACCTCAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATGGACAG  
 GTTATCAACGAAACTCTCAGCATACGATGACCTTGAA  
>ZHX1  
 GAACAAACAATAATGATCTGACTTTGATGGTAGTTTGTAAAGAGGAGAATGCAAGAGCAAGCA  
 GAATCTACAGAAGTTCTTCGGGAATATCTATCAGTAAAACCTTATCATGAAAATGATGAAA  
 AATAAAGTGGAAAATAACCGGATTGCAGTTCATATAACTCAGTTGAGGACGTTCTGAAAGAGAAA  
 GAGAATGAAATCAAACCAGACCGTGAAGAAATTGAGAAAATCCAAGTTCTCAGCTCTGAATCT  
 AATACAAGTACTCCATTGTAACAGAATACATCCAAGTACTGCCAGCACGGTAGTGACACCAGCA  
 GCAGTTCTCCTGGATTGCCACAGGTGATAACTGCTGATCTGCTCAGCAGAATTCTAATTGATT  
 CCCAAAGTCTTAATCCCTGTTAATAGCATTCCCACCTACAATGCTGCATTGGATAACAATCCCCCT  
 TTACTTAAACACCTACAACAAGTTCCCTTACCCAAACAATGTCAGAAATTACAGTTCTGCTCAA  
 GCAAAATATAACAGAGGAACAGATCAAGATATGGTTTCAGCCAACTGTTAAACATGGTGTAGT  
 TGGACTCCGGAGGAAGTAGAGGAGGAAGAAGGAAACATTCAATGGAACAGTGCATACTGTACCT  
 CAGACCATAACTGTTATTCTACACACATTCCACAGGGAGTAATGGTTACCATCTATTTACAG  
 ACATGCCAAATAGTTGGTCAGCCTGGTCTGGCTTACTCAAGTGGCTGGAAACAAACACCTGCCA  
 GTTACAGCACCTATAGCCTGACAGTGGCAGGGTCCAAGTCAAAATAATATACAGAAAAGTCAG  
 GTACCTGCTGTCAGCCTACTGCAGAAACAAAGCCAGAACAGCAGCAGTTCCAACCTCTCAAAGT  
 GTCAAACATGAAACTGCCATTGGTAACCCCTGATTTCATTGGCATTCGGGAAACAAAAAGACAAAAGAG  
 CAACTGGCAGAATTAAAAGTTAGCTACCTTAAACATGAGTTCCCATGATTCAAGAAATTATCAGA  
 CTTATGAAAATAACAGGCTGACGAAAGGAGAGATTTAAAGGTTAGTGCACACAAGGTACAAC  
 CAGAGAAATTCAAAGAGTAATCAGTGCCTACATCTCAACAAATGATCCTCTACCACCATTTATA  
 GACTCCAGTGTGAAACCAACCGGAATCCCCAACTGTTGGTACTGCACAGCCTAAGCAATCCTGGAAT  
 CCTTTCTGACTTTACTCCCCAAAGTTAAAGAGAAAACCTGCAGAGCAGCTCGTGTCTTCAG  
 GCAAGTTCTCAACAGCTGTACTTACAGATGAAGAATTAAAGGTTAAGGGCACAAACCAA  
 CTTACCGAGAGAAATCGATGCTGGTTACAGAGAAGAAGAAATCAAAGCTTAAAGGAAGAG  
 AAAATGAAAATAGATGAAAGTAATGCAGGTAGTTCCAAAGAAGAAGCTGGAGAAAATCTCTGCA  
 GATGAATCTGGTGCACCTAAGTCAGGGAGTACAGGCAAGATATGTTAAAAAAACACCTGAGCAGCTG  
 CACATGCTTAAGAGTGCATTGTCGGACACAGTGGCCATCACCAGAACAGAGTATGACAAGTTGGCC  
 AAAGAAAAGCGGGCTTGCTAGAACAGACATAGTTAGTTGGTTGGGACACCCGTTATGCTTGGAAAG  
 AATGGAAAATGAAATGGTACTACTATCAGACGCCAATTCAAGTAGTATGAATGGTCTGTCT  
 TCCCTTAGGAAAAGAGGGAGAGGGAGACCCAAAGGACGGGAAGAGGAAGACCGCGTGGCGGCGCT  
 AGAGGAAGCAAAGAATTAAACAACTGGGACAGGGGACCATCACTCATAAAATTAAACTGGAACT  
 GCAATACTTAAGGATTATTACCTGAAGCACAAGTTCTTAATGAGCAAGACCTTGATGAACCTGTT  
 AACAAATCACATATGGCTATGAGCAGGTAGAGAGTGGTTGCAGAAAGACAGAGAAGATCAGAA  
 TTAGGTATAGAATTATTGAGGAAAATGAGGAGGAAGATGAAGTTATTGATGACCAGGAAGAGGAT  
 GAAGAAGAAACAGATGATGACACTTGGGAACCTCCACGACATGTGAAACCGGAAGCTGTCTAAA  
 TCAGATGAC

Figure 6 (continued)

&gt;ZNF33B

TGTTATGAATGTGGAAAACCTTCTGCTGAAGTCAGACCTCACAAATAACATCAGAGAACGCACACA  
GGGGAGAACCCTTGCATGCTGAATGTGGAAATTCTTAGCCATAAGTCACCCCTCTCAA  
CATTATAGAACACACACGGGGAGAAACCTATGAATGTCATGAATGTGGAAAAATCTTTACAAT  
AAATCATACCTAACAAAACATAATAGAACACATACAGGGGAGAAACCTATGAATGTAATGAATGT  
GGAAAAACCTTCTGCCAGAAGTCACAACACTCACTCAGCATCAGAGAATTACATAGGGGAGAAACCC  
TATGAATGTAATGAGTGTGGAAAAGCTTCTGCCATAAGTCAGCTCTAATTGTACATCAGAGAAC  
CATACACAAAGAAAAGCCTTATAAATGTAACGAATGTGGAAAATCTTCTGTGTGAAGTCAGGACTT  
ATTTTACATGAGAGAAAGCACACGGGGAGAAACCTATGAATGCAATGAATGTGGAAATCCTTC  
AGTCACAAATCATCACTCACAGTACATTACAGGGCTCACACAGGAGAGAAATCTTGTCAAGTGTAA  
GAATGTGGAAAATCTTTACCGTAAATCAGACCTTGCTAACATCAGAGATCACATACAGGGAA  
AAGCCCTATGAATGTAACACATGCAGGAAAACCTTCTCAAAAGTCAAATCTCATTGTACATCAG  
AGAACACACATAGGAGAAAACCTTATGAA

Figure 6 (continued)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**